Digitized by the Internet Archive
in 2009 with funding from
Lyrasis Members and Sloan Foundation

http://www.archive.org/details/notesonphysics00snow

Edward J. van Liew

808 University Av

adison

Wiscons

1914-1915

West Virginia University Libraries
QC 21.S65 1914 hsmo
Notes on physics,
3 0802 001028975 4

History Room
QC
21
.S65
1914
Snow, Benjamin Warner
Notes on physics.

DO NOT CIRCULATE

QC
21
S6
1914

NOTES ON PHYSICS

BY

BENJAMIN WARNER SNOW

PROFESSOR OF PHYSICS IN THE
UNIVERSITY OF WISCONSIN

MADISON, WISCONSIN
1914

Copyright, 1914
By Benjamin W. Snow

NOTES ON PHYSICS

CHAPTER I.

MATTER AND ITS CONSTITUTION.

1. **The Study of a Natural Science.** The science of physics has for its province a study of the phenomena that occur in the world of nature about us, and the endeavor to understand more fully the laws according to which these phenomena take place. The human mind, though filled with admiration at the manifestations of nature which abound on every hand, is not content to receive impressions of these various occurrences simply as an assemblage of separate and disconnected events. It seeks, rather, to discover, if possible, a relation between these phenomena, and to investigate their causes, in order to draw conclusions as to the laws of the universe, in accordance with which these results are produced. Moreover, when we contemplate the harmony, the beauty, and the uniformity of these laws, we learn very soon that there is no change and no deviation in their operation, and that causes, which at one time produce certain effects, will, under similar conditions and at all other times, produce the same effects.

During the study of a natural science, sometimes for the purpose of illustration, and sometimes in order that a conclusion may be drawn or the relation of one law to another inferred, it is often necessary to observe the effects which the operation of a particular law will produce, and this is accomplished by an experiment. It must be borne constantly in mind, however, that without exception the experimenter simply supplies certain conditions which render possible the operation of this law; and the effects of the law then follow as a necessary and inevitable result. The unity and the constancy and the invariability of Nature are what give to the student his faith and his belief in the study of Nature's laws; and the effort to see farther and more clearly into obscure phenomena, and to understand more fully the causes that underlie observed effects, as well as to comprehend more intelligently the laws according to which these effects are accomplished, is the inspiration of the scientist.

2. **Matter.** When we observe the various objects in the world about us, and recognize that these objects are *material,* or are composed of *matter,* questions then naturally arise in our minds as to how this matter exists, what its structure may be, and of what it itself is composed. The attempt to answer these questions has interested men of all ages, but even now we are forced to admit that the nature of matter is so imperfectly known that we are as yet unable to tell what matter really is. The evidence of our senses, however, discloses the fact that matter has objective existence in the three-dimensional space in which we live; and so we may use this property of space-occupation as a provisional means of recognizing the existence of matter, and define *matter to be anything that fills or occupies space.* A limited portion of matter is called a *body,* and an extremely minute quantity is known as a *particle.*

Our understanding of this subject is further confused by the fact that matter exists in different *states,* or *conditions*—the *solid,* the *liquid,* and the *gaseous*—and is of many different *kinds,* while a given space may contain a greater or less *quantity,* depending in part upon the *quality* of the matter enclosed by that space. Chemistry, indeed, recognizes about eighty different primary kinds of matter called *the elements,* and it is from combinations of two or more of these that matter of all kinds is made.

3. **Divisibility.** Because of the property of manifold division which is common to matter of all kinds, and which is known as divisibility, matter may be separated into smaller and smaller portions. So far is it possible to carry the subdivision of matter, that we are wholly unable to comprehend the size of the minute

particles resulting from the further partition of a small portion of many substances. Thus, it has been calculated that the one-hundred-millionth part of a grain of sodium will give color to the flame of a bunsen burner or of an alcohol lamp; a thimbleful of one of the aniline dye-stuffs is quite sufficient to impart a tinge of color to the water of a large lake; while a single grain of musk, after perfuming the air continuously for many years, will still emit a perceptible odor. From these illustrations, it is seen that the individual particles of which bodies are composed must be exceedingly minute, and that their number, likewise, must be enormously great.

4. The Molecules. Although matter may thus be divided and subdivided until the resulting particles are almost infinitely small, there is a limit beyond which it is impossible to carry this division without destroying the characteristics of the original substance. The smallest particles of matter that can exist and still retain the properties of the larger quantity are called the *molecules*. If we were to divide the molecules, we should cease to have the particular kind of substance with which we are dealing; and so we may define the *molecules* to be *the smallest particles of matter that can exist and still preserve the distinguishing qualities of the undivided whole.* Since the molecules are far too minute to be detected, even with the aid of the most powerful microscope, we understand but little of the nature of these particles, and this scanty knowledge is obtained only by inference and not by direct observation.

From many of the phenomena connected with matter, we are led to believe, however, that the molecules are not at rest, but are in a state of *rapid and irregular motion*. So great is their number, and so violent their motion, that they are continually bombarding one another and rebounding from such collisions only to encounter other molecules which are moving no less energetically than they.

That the molecules of a gas are in this state of perpetual motion may be shown by inverting a tall jar of ammonia gas over one of hydrochloric acid gas, the two gases being separated temporarily by a plate of glass. When all currents in these gases have ceased and the gases are perfectly quiescent, the glass may be withdrawn. It will then be seen, by the white cloud which is formed by the union of these two gases, that the molecules of the lighter ammonia gas will force their way downward into the heavier hydrochloric acid gas, while the molecules of the latter will immediately wander upward into the lighter ammonia gas. Therefore, the two gases are quickly and completely mingled, because the molecules of both are continually in motion.

5. The Pores. While it is true that the molecules are minute beyond our comprehension, and the quantities of matter with which we have to deal are simply aggregations of these molecules, there are certain phenomena, particularly in the domain of heat, which lead us to believe that the molecules of a body are not in contact with one another. Indeed this isolation of the molecules might be inferred from the assumption that the molecules, in their continuous and irregular motion, are able to deliver blows upon one another. It is thought, moreover, that vacant spaces exist between these particles, and that these spaces are very large in comparison with the dimensions of the molecules themselves. These inter-molecular spaces are known as the *pores;* and it is believed that pores are to be found between the molecules of all substances, although in some kinds of matter they are smaller than in others. When a body expands with a rise of temperature, it is, in general, assumed that the molecules are farther apart, and that the pores, accordingly, are somewhat larger, than is the case at lower temperatures.

Sir Francis Bacon made a hollow sphere of lead, which he filled with water and tightly closed. After partially flattening the sphere, whereby the interior volume was somewhat diminished, he found that the water had been forced through the pores of the lead, and stood in drops upon the outside. The members of the Florentine Academy likewise constructed a hollow sphere of silver, heavily gilded, which they treated in a similar manner, with the result that the water was forced through the pores of the silver and covered the sphere in tiny beads, like perspiration. From these experiments it is evident, not only that particles of water had passed through the spaces between the molecules of the lead, the silver, and the gold, but also that the molecules of water are smaller than are the pores of these metals.

The existence of pores even in a liquid may be demonstrated by filling a long glass tube half full of water, then carefully filling the remainder with alcohol, and observing, by a mark on the tube, the height to which these liquids rise. If the tube is now shaken until the liquids are thoroughly mingled, the surface will be found considerably below the former mark, showing that the volume of the mixture is somewhat smaller than the sum of the volumes of the water and the alcohol. This contraction of volume is explained, in part at least, by assuming that the liquids have *interpenetrated,* that is, that the molecules of one have to some extent entered the pores of the other.

Charcoal has the power of absorbing, and also of condensing, gases within its pores, as may be illustrated by placing a piece of freshly ignited charcoal in a tube of ammonia gas inverted over mercury. The molecules of the gas enter the pores of the charcoal and there undergo a condensation, so that a portion of the gas disappears from the tube, and the mercury rises to take its place.

6. **Molecular Attractions.** Because of the existence of the pores, the molecules of a body are thought to be separated by relatively very large spaces, each molecule being in the center of a vacant region surrounded by other molecules which it never touches. Although not in actual contact, these molecules exert upon one another a powerful attraction, of such nature that the distance between the molecules does not exceed a definite amount, nor does it fall below a certain limit. It is due, indeed, to this inter-molecular attraction that the particles of a body are held together, and that matter possesses the property known as *tenacity.* In some bodies, such as steel, the particles are united more firmly than in others, such as lead or copper, and so the tenacity of steel is much greater than is that of these other substances.

7. **Cohesion and Adhesion.** If a body is homogeneous, that is, if it consists of one kind of matter only, the attraction between the molecules is known as *cohesion.* If, however, it is

heterogeneous, or is composed of matter of different kinds, the molecular attraction is known as *adhesion.* A piece of steel, for instance, maintains its rigidity because of the *cohesion* of its molecules; a piece of steel cemented to a piece of wood, however, clings to the wood by the *adhesion* of the cement. The force of *cohesion,* therefore, holds together *molecules of the same kind,* while the force of *adhesion* unites *molecules of different kinds.*

8. **Spheres of Molecular Activity.** These molecular forces are found to be relatively very great, and to increase enormously in magnitude as the distance between the molecules becomes smaller. Were it not, therefore, for the action of other forces, these attractions would probably cause the molecules to come into actual contact. Other forces become manifest, however, when the attempt is made to crowd the molecules closer together; and these latter forces are so constituted that they apparently cause the molecules to repel one another when the intervening distance becomes less than a certain limiting amount. Accordingly, we may think of one molecule as exerting its influence upon adjacent molecules across a perfectly definite distance. This distance, which extends about each molecule in all directions, Fig. 1, is known as the *sphere of molecular activity.* If this assumption be true, it follows that the molecules may approach one another until their spheres of activity are tangent; and it is only with great difficulty that they may be forced nearer together or carried farther apart than the diameter of their spheres of activity.

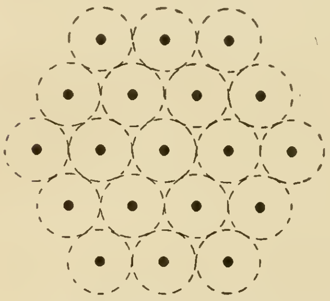

Fig. 1. Molecules and spheres of molecular activity.

9. **Nature of Elasticity.** It is proved by experience, nevertheless, that the action of an external force will slightly vary the distance between the molecules, and so, probably, change the spheres of activity temporarily into ovals. The distortion thus

produced at once calls into play molecular forces which attempt to restore the molecules to their original positions, and enable the ovals of activity again to assume the spherical form. This property of matter, whereby a body tends to resume its original shape and size after it has been distorted by a force, is known as *elasticity*. It is thus seen that the elasticity of matter depends entirely upon the action of the inter-molecular forces in endeavoring to maintain a constant distance between the molecules.

10. The Atoms. The atoms are the minute particles of matter of which the molecules are composed, and are very small as compared with the size of the molecules themselves. In general, the atoms unite with one another, in structures more or less complex, to form the molecules; the number, kind, and arrangement of the atoms being perfectly definite in a molecule of a particular substance. The atoms thus grouped in a molecule are held together by remarkable atomic forces which prevent these particles from separating from one another, and probably, also, from coming nearer together. It is possible that the atoms are relatively as far apart in the molecules as the molecules are distant from one another in masses of matter. Moreover, certain phenomena in physics lead to the conclusion that the atoms in a molecule are not at rest, but are in a state of continual motion, either forming systems, in which certain atoms revolve about other atoms, or vibrating about a common center, like the parts of an elastic hoop, Fig. 2.

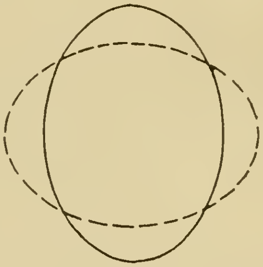

FIG. 2. The possible vibration of the atoms in a molecule.

11. The Chemical Bonds. There are as many different *kinds* of atoms as there are chemical elements, and all differ from one another as regards their composition, their relative weights, and their ability to combine with other atoms to form molecules. Whatever their constitution may be, most atoms have at least one very active point, by means of which they are able to attach themselves to other atoms. These points of activity are called the *chemical bonds;* and it is a curious fact that, while nearly all atoms have at least one bond, some have two, others three, and so on, a few of the elements having a number as high as seven. In recent years, however, certain inert gases have been discovered in which this bond is either feeble or is absent altogether, and in these gases the single atom constitutes the molecule. It has also long been known that the molecules of mercury, cadmium and potassium when in the state of vapor are likewise monatomic. In all other substances, this exceedingly active chemical bond, with at least one of which every atom is equipped, renders it impossible for atoms to exist by themselves; and so, except in the rare instances mentioned above, two atoms at the very least unite to form a molecule, the atoms being joined to one another by their chemical bonds.

Thus, in hydrogen, Fig. 3, one *atom* of hydrogen, *H*, unites with another *atom* of hydrogen, *H*, to form a *molecule* of hydrogen, H_2, each atom being joined to the other by the single chemical bond which this element possesses. So, also, in the case of common salt, one *atom*, of sodium, *Na*, unites by its single bond with one *atom* of chlorine, *Cl*, which also has a single bond, the two atoms together forming a *molecule* of sodium chloride, *NaCl*, or common salt. In water, likewise, one *atom* of hydrogen, *H*, unites with *each* of the two bonds of an *atom* of oxygen, *O*, to form a *molecule* of water, H_2O, which thus contains three atoms. The number of atoms in a molecule of complicated chemical compounds may amount to several hundred.

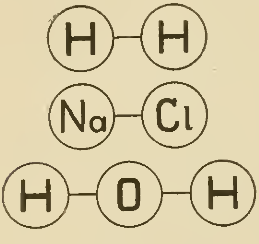

FIG. 3. Molecules formed by the union of atoms.

12. The Distinction between a Physical and a Chemical Change. If the molecule is by any means destroyed, and the atoms that formed certain groups within it are thereby forced to separate, these atoms will at once form other and different groups because of the activity of their chemical bonds. The molecules resulting from this new assemblage of atoms will therefore have properties unlike those of the original material, and an entirely different chemical substance will be formed.

The science of physics has to deal almost exclusively with matter and the phenomena connected with matter so long as the molecule is not destroyed. Chemistry, on the other hand, has for its province the changes that occur in matter and the phenomena attending these changes, when the molecule is broken up. A physical change performed upon matter may alter the size, the shape, the motion, the hardness, the tenacity, the elasticity, and many other properties of this matter, but will leave the molecule intact. A chemical change performed upon the same matter will alter the constitution of the molecule and so change completely the character of the substance. After a chemical change has occurred, the original material will no longer exist, and in its place will be formed one or more different substances consisting of the same atoms, but forming other groups within the molecule.

A beautiful illustration of the difference between a physical and a chemical change may be shown by the experiment of pouring strong sulphuric acid upon a solution of common sugar. Sugar is a substance which has 45 atoms in its molecule, 12 of carbon, C, 22 of hydrogen, H, and 11 of oxygen, O; and this compound is chemically expressed by the formula $C_{12}H_{22}O_{11}$. We may break the sugar into lumps, we may grind it to a powder, we may even dissolve it in water, but it still tastes sweet, and retains in all other respects the properties and characteristics of sugar. It has undergone therefore only certain physical changes, for its molecule has not been destroyed.

The hydrogen and the oxygen, as will be observed, exist in the sugar in the proportions in which these elements are combined in water, that is, there are twice as many atoms of the former element as of the latter. If it were not for the chemical bonds, uniting the hydrogen and the oxygen to the carbon in the molecules of sugar, the carbon atoms would join with one another to form molecules of carbon, and the hydrogen and the oxygen atoms would unite to form molecules of water. The sulphuric acid has an exceedingly strong chemical affinity for water, and will absorb it from compounds wherever it exists. Because of this strong affinity, the sulphuric acid, when poured into the sugar solution, breaks up the molecules of sugar, and withdraws from them the atoms of hydrogen and oxygen to form molecules of water, with which it then unites. The carbon atoms thus left free, combine with one another to form molecules of carbon, or charcoal, and this is seen to be produced in large quantities as the result of the experiment. In this instance the molecule of sugar has been destroyed, the sugar has ceased to exist as sugar, the atoms of carbon, formerly united with the atoms of hydrogen and oxygen to form sugar, are now united with one another to form molecules of carbon, or charcoal, and a chemical change has been accomplished.

13. The Electrons. The word atom signifies "that which cannot be cut." It formerly indicated a particle of matter so small as to be incapable of further division, and was thought to have existence, not, indeed, by itself, but only in union with other atoms to form a molecule. In accordance with this idea, the atoms have been, and are still, defined as the smallest particles of matter that *can enter into chemical composition.* They are thus distinct from the molecules, which are regarded as the smallest particles of matter that may be conceived to *exist by themselves* and still preserve the identity of the original substance.

Within the last few years, however, certain physical processes have served to show that the atoms themselves are far from simple in their structure, and that even these small particles are composed, in part at least, of other still smaller particles no larger than the $\frac{1}{1700}$ part of the hydrogen atom, and are known as the electrons, or the corpuscles. The simplest known atom, namely, that of hydrogen, is supposed to contain not less than three nor more than five of these electrons, while every other atom has a number greater than this in proportion to its atomic weight.

The evidence of scientific investigation points likewise to the conclusion that these electrons are not at rest within the atom, but are *moving,* possibly in tiny orbits, at a rate which is *but little less than the velocity of light.* It has disclosed the fact, also, that, under certain conditions, an electron may escape from its parent atom, and, while still retaining its enormous velocity, may exist as *a free and independent particle among the molecules of matter,* bombarding these in its rapid motion, and being in turn bombarded by them. From this it is seen that the electrons may exist either in rapid motion within the atom of which they form an essential part, or as individual particles moving with enormous velocities among the molecules of ordinary matter. In the latter condition these particles are known as the *free* electrons.

Inasmuch as it is possible for certain of the electrons thus to be disengaged from the atoms and to exist freely by themselves, a study may be made of the physical properties and characteristics which these small bodies possess. It is then found that they are *all alike,* even when obtained from substances differing widely from one another in character. From this it it evident that the electrons form a *common constituent of the atoms of all kinds of matter.* Moreover, the electrons are always *negatively electrified,* they always carry precisely the *same electrical charge,* and always have an *apparent "mass"* which is dependent upon the *velocity* with which they move. Indeed, it is thought that the electrons may not be composed of matter in the ordinary sense, but that they may be simply particles of negative electricity which have apparent mass only because of their enormous velocity.

The important part which these electrons are believed to play in the phenomena of heat, light, and electricity, and their significance in the newer theory that matter may possibly be electrical in structure or constitution, will be discussed in later chapters.

CHAPTER II.

THE METRIC SYSTEM.

14. The Fundamental Quantities. In the pursuit of the study of physics, it is found impossible to investigate the relation which one phenomenon bears to another, or to predict the results which will follow from a given cause, without knowing accurately the magnitudes of the objects or quantities which are thus involved. This knowledge is therefore of the greatest importance to the physicist, and may be obtained only by means of exact measurements.

Before a measurement of any kind may be made, it is necessary first to adopt a *standard,* which must possess two essential characteristics: it must be of the same *kind,* or *nature,* as the object or quantity to be measured, and it must be *unchangeable.* If this standard is of suitable magnitude to be used conveniently, it itself may then be taken also as the *unit* in terms of which the measurement is to be expressed. If, however, the standard should be inconveniently large or unduly small, a fractional part or a multiple of its magnitude may be taken as the unit, and the measurement then shows how many times the quantity to be measured is larger or smaller than the unit so chosen.

Observation and investigation have shown that, by a proper selection of units, the measurement of practically all physical quantities may be expressed in terms of *three* so-called *fundamental quantities:* namely, *length,* or linear distance, *mass,* or quantity of matter, and *time.* These three quantities are called the *fundamental quantities,* and the units of these quantities are known as the *fundamental units,* because measurements of all other quantities in physics may be expressed in terms of the units of these three particular quantities

15. The Standard of Length—The Meter. The standard of length is the linear distance at the temperature of melting ice between two fine, parallel lines ruled near the ends of a bar of platinum-iridium deposited in the International Bureau of Weights and Measures, at Sèvres, a suburb of Paris. The perpendicular distance between these two parallel lines on this particular bar is defined as the *international meter prototype,* and is the standard of length universally adopted in all scientific measurements. In order that this standard may be accessible in all countries, copies of the international meter prototype are deposited at the capitals of the various civilized nations of the globe. The cross-section of this bar and of its copies resembles somewhat the letter X, Fig. 4, since this form offers the greatest rigidity for a given amount of material. The graduations which determine the length of the meter are ruled at *A* upon what is known as the plane of the neutral axis, in order as far as possible to avoid distortion by flexure.

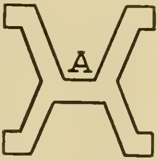

FIG. 4. Cross-section of the International Meter Prototype, exact size.

With the greatest accuracy that could be attained in making linear measurements, the distance between the two parallel lines on the international meter prototype was made equal to the distance between the two ends of a particular bar of platinum deposited in the Archives of Paris, which was known as the *meter of the Archives.* This platinum bar had been constructed to represent the ten-millionth part of the distance from the north pole of the earth to the equator on the meridian passing through Paris, and was adopted as the standard of length in the hope that it might furnish an invariable standard, which, by a re-measurement of the earth, could at any time be reproduced if the original were lost.

It was discovered later, however, that slight errors had been made in the measurement of the earth, and that the length of this bar was about one four-hundred-thousandth part of an inch too short, so that it was not an absolute standard and did not represent the ten-millionth part of the quadrant of the earth, this distance now being known to be about 10,000,880 meters. From this it follows that the international meter prototype, which was made as nearly as possible equal in length to the meter of the Archives, is, after all, only a purely arbitrary standard.

If this bar should be lost or destroyed, the meter could be reproduced, at best only imperfectly, from the various existing copies. The need of an absolute and invariable standard therefore still exists; and it is to be hoped that at some future time a standard of length may be adopted which will depend upon some natural object or occurrence, and which may be reproduced with absolute accuracy, if the original should be lost or injured. The suggestion made by Michelson, that the wave length of red cadmium light be adopted as the standard of length, is a fruitful one, for this distance, though small, would be an absolute and unvarying standard, and capable of reproduction by any body at any time. The means by which this distance could be used as a standard of length will be discussed in a future chapter.

16. Subdivisions of the Meter. Notwithstanding the arbitrary length of the meter, a great advantage has been derived from the adoption of the system of weights and measures, known as the metric system, which is based upon the meter as the standard of length. The advantage which the metric system possesses over all others, and which has brought it into universal scientific use, is the fact that it is a decimal system throughout, the subdivisions of the standards being indicated by Latin prefixes, and the multiples by Greek prefixes. Thus the meter (m) is divided into ten parts each called a decimeter (dm); the decimeter is divided into ten parts, each called a centimeter (cm); the centimeter is divided into ten parts, each called a millimeter (mm). One thousand meters, which is the only multiple of the meter in general use in physics, is called a kilometer (km). Accordingly,

1 meter = 10 decimeters = 100 centimeters = 1000 millimeters.
1000 meters = 1 kilometer.

In terms of the ordinary English units of measurement, the meter has a length of 39.37 inches, the decimeter is nearly 4 inches, Fig. 5; the centimeter is a little less than half an inch, and the millimeter is about one twenty-fifth part of an inch. A kilometer is nearly five-eighths of a mile.

17. The Units of Length, Area, and Volume. While the meter is the *standard* of length, the *unit* of length which is universally adopted in all physical measurements is the one-hundredth part of the meter, or the *centimeter*. In physics, therefore, all linear distances are expressed in centimeters.

As in the English system, so also in the metric, the units of area and of volume are derived from the unit of length. Thus, a *square centimeter* is a surface having an area equal to that of a square one centimeter on a side; a *square meter* is a surface having an area equal to that of a square, one meter on a side, etc. A convenient unit of *volume* which is of frequent use in physics is the volume of a cube, each of whose sides is one centimeter in length, and known as the *cubic centimeter* (cc). Another unit of volume frequently used is the *cubic decimeter* (cd), which is the volume of a cube 1 decimeter, or 10 centimeters, on a side. The volume of one cubic decimeter therefore, is the same as that of 1000 cubic centimeters, and this volume is practically identical with that known as the *liter*. The liter is very nearly equal in volume to the ordinary quart.

1 Decimeter divided into Centimeters and Millimeters.

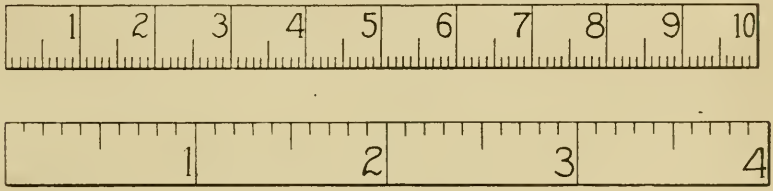

Fig. 5. Inches and Tenths.

18. The Standard of Mass. By the *mass* of a body is meant the quantity of matter the body contains.

The standard of mass, in terms of which the amount, or quantity, of matter in all bodies is measured, is a cylinder of platinum-iridium, deposited in the International Bureau of Weights and Measures at Sèvres, and known as the *international kilogram prototype*. This standard was made to represent as nearly as possible the mass of the *kilogram of the Archives,* which was a cylinder of platinum, adopted in 1799 as the standard of mass, and deposited in the Archives of Paris. In order to facilitate measurements throughout the world, copies of the international kilogram prototype have been distributed among the various nations.

One of the most beautiful features of the metric system, as this was conceived in the minds of its founders, was the possibility, not only of securing an absolute and invariable standard of mass, but also of making this standard depend for its magnitude upon the unit of length. It was proposed that the standard of mass, which was to be called the kilogram, should be a cylinder of platinum equal in mass to that of 1 cubic decimeter, or 1000 cubic centimeters, of pure water at the temperature of its greatest density; and the kilogram of the Archives was constructed to represent such a mass. Subsequent accurate measurements, however, have shown that the kilogram was made slightly larger than was intended, and that its mass was really equal to that of 1.00003 cubic decimeters of pure water. From this lack of exact agreement, it is seen that the kilogram of the Archives was not an absolute standard; and therefore, the international kilogram prototype, which is derived from it, is likewise not an absolute standard, but is unfortunately one which is purely arbitrary.

19. The Unit of Mass, the Gram. The relatively large mass of the kilogram made is desirable that a smaller mass should be chosen as the *unit;* and the one-thousandth part of the kilogram was taken as the *unit of mass,* and was called the *gram.*

The subdivisions and multiples of the gram are decimal, and, like those of the meter, are indicated by the Latin and the Greek prefixes respectively. Thus, one gram (g) is divided into ten parts, each called a decigram (dg); one decigram is divided into ten parts, each called a centigram (cg); one centigram is divided into ten parts, each called a milligram (mg). The only multiple of the gram in general use in physics is that by one thousand, and this, as shown above, is the kilogram (kg). Accordingly,

1 gram = 10 decigrams = 100 centigrams = 1000 milligrams.
1000 grams = 1 kilogram.

The kilogram is very nearly equal to 2.2 pounds.

If the mass of the kilogram had been made *exactly* equal to that of 1 cubic decimeter of pure water, the mass of the gram would have been *exactly* that of 1 cubic centimeter of pure water, and the unit of capacity, which is the *liter,* and which is defined as the volume occupied by the mass of 1 kilogram of pure water at the temperature of its greatest density, would have been 1 cubic decimeter, or 1000 cubic centimeters. Inasmuch, however, as the mass of the kilogram is so very nearly equal to its intended value, it is allowable and usual, except in measurements of the utmost accuracy, to regard the *kilogram* as the mass of *1 cubic decimeter, or 1000 cubic centimeters,* of pure water at the temperature of its maximum density, and the *gram* as the mass of *1 cubic centimeter* of pure water, likewise at the temperature of its maximum density. The liter also may be taken, practically without error, to be the volume of *1 cubic decimeter, or 1000 cubic centimeters.*

20. The Standard of Time and the Unit of Time. The period that has been adopted as the *standard of time* is the interval known as the *mean solar day.* Owing to the fact that the earth revolves in an elliptical orbit about the sun, and that its axis is not perpendicular to the plane of its orbit, the solar day, that is, the time from noon to noon, or the interval that elapses between two successive passages of the sun across the meridian, differs from day to day. The mean, or average, of these solar days for an entire year is known as the *mean solar day.*

The mean solar day is divided into 24 parts, each called an hour (h); the hour is divided into 60 parts, each called a minute (min); the minute is divided into 60 parts, each called a second (sec). Except in certain astronomical measurements, the *unit* of time that has been universally adopted is the *mean solar second,* which is therefore $\frac{1}{24\times60\times60}$ or $\frac{1}{86400}$ part of the mean solar day. This is the second that is indicated by our ordinary timepieces.

21. The Equation of Time. Since the actual solar days are of variable length, while our clocks and watches keep mean solar time, it follows that only at certain periods during the year does the actual solar time agree with the mean solar time. Indeed, there are but four days: namely, April 15, June 15, August 31, and December 24, when the sun is on the meridian at the time which we call twelve o'clock. The deviation from this coincidence may be as great as 14 min 29 sec, which occurs on February 11, when our mean-time clock is faster than the sun, or 16 min 20 sec, which occurs on November 1, when our mean-time clock is slower than the sun. Accordingly, in the determination of time by the sun-dial, it is necessary always to apply a correction, known as the *equation of time,* in order to reduce the true solar time, which is there indicated, to the mean solar time, by which our lives are regulated.

22. **Derived Quantities and Derived Units—The C. G. S. System of Measurement.** All quantities except the three fundamental quantities are known as *derived quantities,* and all units except the three fundamental units are known as *derived units.* Thus, an *area,* being derived from two dimensions of *length,* is a derived quantity, and the unit of area, the square centimeter, being derived from the unit of length, is a derived unit.

The *dimensions* of a quantity indicate the number of times the fundamental quantities enter as factors into the expression of that quantity. Thus, an area, being a length multiplied by a length, has for its dimensions $[L^2]$, where $[L]$ is used to denote the *dimension* of length.

Since all physical quantities may be expressed in terms of one or more of the fundamental quantities, *length, mass,* and *time,* and since the units of these quantities are the *centimeter,* the *gram,* and the *second,* respectively, the system of units upon which all measurements in physics are ultimately based is known as the Centimeter-Gram-Second system, or, more briefly, as the C. G. S. system.

23. **The Absolute System of Units.** An absolute system of units is one which is based upon fundamental units that are not only invariable but also independent of locality, and in which the derived units bear the simplest possible relation to the funda-

mental units. Thus, in an absolute system, the unit of area is the square of the unit of length, the unit of volume is the cube of the unit of length, etc., no coefficients or factors of proportionality being used. From this it is seen that the C. G. S. system is an absolute system, as is also the British system, in which the fundamental units are the foot, the pound, and the second. A system, however, in which one of the fundamental units is based upon the *weight* of a body—the so-called gravitational system, for instance—is not an absolute system; for weight varies from place to place upon the earth's surface.

MECHANICS

SOLIDS

CHAPTER III.

CHARACTERISTICS AND LAWS OF MOTION.

24. Position. By the position of a particle in space is meant the location of that particle with respect to a selected reference system which is at rest. To obtain such a system, it is usual to assume three straight lines, or axes, known as the *co-ordinate axes,* at right angles to one another, and meeting at a point called the *origin,* like the three straight lines which meet at a point in the corner of a room. Thus, Fig. 6, O may be considered as the origin, and the lines OX, OY, and OZ as the co-ordinate axes, the latter being commonly called the axes of X, Y, and Z, respectively. The three planes, XY, YZ, and XZ, which these axes determine, corresponding to the floor and the two adjacent walls of a room, are known as the *co-ordinate planes.* When a reference system of this kind has been chosen, it is possible to define the position of a particle, such as P, by three distances, x, y, and z, measured from the origin along the co-ordinate axes. It is possible also to determine the position of P by the length of the line OP which connects this

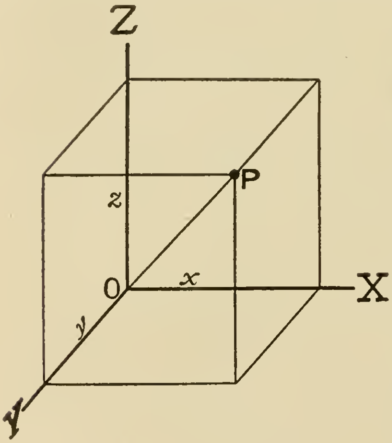

FIG. 6. Axes and co-ordinates.

particle with the origin, and by the angles which this line makes with any two of the co-ordinate axes. The quantities which thus specify the position of a point or particle with respect to the reference system, such as the length of the lines x, y, and z, are called the *co-ordinates* of that point or particle.

25. Motion. When a change occurs in any one or in all of the three quantities which determine the position of a particle, the particle is said to *move*. The *motion* of a particle may therefore be defined as a *change of position* of that particle with respect to the assumed reference system or other particles, the relative positions of which do not change.

Thus an object may be transferred from one part of a table to another, or a person may move from one room to another, while the table and the other objects in the room remain at rest. These other objects, however, while they may not change their positions with reference to the walls of the room, or even to the earth, are nevertheless in motion, for the earth itself is in motion about its axis and around the sun, and carries with it all bodies upon its surface. Accordingly, the objects which were supposed to be at rest upon the earth are indeed at rest with reference to this body, but are really in motion with respect to the sun. The sun, moreover, and also the other so-called fixed stars are know to be in motion through space, for they are observed to change their positions with respect to one another. This being so, all motion and all rest in the universe must be considered simply as relative and not as absolute, for there is no known body that does not change its position with respect to some other body. "It is, therefore, unscientific", says Maxwell, "to distinguish between rest and motion, as between two different states of a body in itself, since it is impossible to speak of a body being at rest or in motion except with reference, expressed or implied, to some other body."

When a body moves, the *path* along which the motion takes place represents the successive positions which the body has occupied. There can be *no discontinuity in motion*, that is, matter cannot disappear from one position and reappear at another without having traversed some intermediate path. When motion has taken place, it is absolutely certain that matter has occupied every point of the path between the initial and final positions of the moving body, however devious this path may be.

Another element of prime importance in the idea of motion is the *time* that is thereby necessarily consumed. Whether a motion is rapid or slow, or whatever its direction may be, *a perfectly finite time* is required for a body to change its position with respect to surrounding objects. Besides the matter which moves when motion takes place, and this will be studied in detail later, there are *three inherent ideas* in the conception of motion, one being the *direction* of motion, another the *length,* or *distance*, over which the body passes, and the third the *time* that is spent in traversing this distance. No motion can take place unless *direction, distance,* and *time* are involved; and the study of motion leads to the consideration of these three quantities, and the relations which they bear to one another.

26. Uniform Motion and Varied Motion. When we observe the different bodies which move in the world about us, we perceive that, while all change their positions, some do so with a *uniform* motion, and others with a motion that is *varied.* A body moves with *uniform motion* when the direction of its motion remains unchanged, and when it passes over equal distances in equal successive intervals of time. On the other hand, a body moves with *varied* motion when either of these conditions changes; that is, when the direction of motion changes, or when the body passes over unequal distances in equal successive intervals of time, or when both of these changes in motion occur together.

27. Uniform Motion.—Velocity. Since in *uniform* motion *equal distances* are passed over in *equal successive intervals of time,* it follows that twice as great a distance will be passed over in two of these equal intervals as in one. Thus, if a body passes over 2 feet in 1 second, it will pass over 4 feet in 2 seconds, 6 feet in 3 seconds, and so on. From this it is evident that the condition of the motion of this body may be expressed by any one of a number of fractions, the numerators of which represent the distances passed over, and the denominators the times spent in passing over these distances. Moreover, the values of these fractions are all equal to one another, and are equivalent to the distance passed over in the unit of time.

When we thus divide the whole distance through which motion has occurred by the time consumed in this motion, we obtain the *rate of motion,* or the *rate of change of position,* and this quantity is termed the *velocity.* Accordingly, *velocity* is defined as the *rate of motion,* that is it is the *ratio* of the *distance passed over* to the *number of units of time spent in passing over this distance,* or it is *the distance passed over in the unit of time.* Thus, from the above example of uniform motion, we may obtain the velocity as follows:

$$\frac{\text{Any length}}{\text{Corresponding time}} = \frac{6 \text{ feet}}{3 \text{ sec}} = \frac{4 \text{ feet}}{2 \text{ sec}} = \frac{2 \text{ feet}}{1 \text{ sec}} = \text{velocity}.$$

from which we know that the body moving with this motion has a velocity of *2 feet per second.* In speaking of a velocity, it is necessary to mention *both the distance* passed over *and the time* spent in passing over this distance.

Moreover, we perceive that throughout this motion the *velocity remains the same,* or *is constant.* Accordingly, a body that moves with *uniform motion* has a *constant velocity;* and, conversely, a body that has a *constant velocity* moves with *uniform motion.*

28. Speed. When we speak only of the distance that may be described in a certain time, without considering the direction in which the distance may be traversed, it is customary to use the word *speed* instead of velocity. Thus, we may speak of the speed of a race horse or the speed of a train, meaning that a certain distance will be traversed in the unit of time, without reference to the particular direction in which this motion takes place. *Speed,* therefore, refers simply to the *numerical part of velocity,* and is *independent of direction.* Accordingly, to have a *constant velocity,* a particle or body must move with *uniform speed in a straight line.*

29. Relations in Uniform Motion. If the motion is uniform, that is, if the direction remains unchanged and the moving body passes over equal distances in equal times, the relation between the quantities involved in velocity may be shown more simply as follows: If l denotes the entire *length*, or distance, passed over with uniform motion, t the number of units of *time* consumed in the motion, and v the *velocity*, then, since

$$\text{velocity} = \frac{\text{length}}{\text{time}},$$

we may write

$$v = \frac{l}{t},$$

from which

$$l = v\, t,$$

or

$$t = \frac{l}{v}.$$

These three expressions may be interpreted as follows: From the first we understand that if a body moves uniformly a distance of l units of length in t units of time, it will move a distance of $\frac{l}{t}$ units of length in 1 unit of time, or that it has a velocity of $\frac{l}{t}$ units of length per unit of time. From the second we learn that the entire distance l, described by a body in t units of time, is the distance v, which it travels in 1 unit of time, multiplied by t, the number of units of time. From the third we see that the time t occupied by a body in traveling uniformly a certain distance is the ratio of the entire distance l that it travels to the distance v that it travels in 1 unit of time.

The second form of this equation

$$l = v\, t,$$

is of particular service and of frequent application in physics, inasmuch as it expresses the distance passed over by a body or particle which moves with uniform motion. Since the distance

l is the product of the time t and the velocity v, this distance may be represented, as in Fig. 7, by the area of a rectangle, in which the base, or t, represents the time of duration of the motion, and the uniform height, or v, is the constant velocity of the body during its motion.

30. The Unit of Velocity. In the expression for the velocity

$$v = \frac{l}{t},$$

if $l = 1$ and $t = 1$, then $v = 1$. The unit of velocity, accordingly, is the velocity of a body that describes, or travels, the unit of length in the unit of time. The unit of length being 1 centimeter and the unit of time 1 second, the *unit of velocity* is the velocity of a body that passes over *1 centimeter of length in 1 second of time.* The body is then said to have a velocity of 1 centimeter per second.

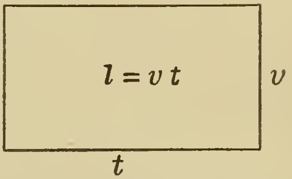

FIG. 7. In uniform motion the distance passed over equals the velocity multiplied by the time.

31. Varied Motion. If the direction of motion changes, however, or if the moving body passes over unequal distances in equal succeeding intervals of time, or if both of these changes in the motion occur together, the motion is no longer uniform but varied. In this case the velocity is also varied, because the distance passed over in one unit of time is not the same as that passed over in another unit of time. When the motion is varied, the velocity changes from instant to instant, and can have a meaning only when we specify the particular moment or position at which it is to be measured. Even when this is done, a variable velocity cannot be expressed by the distance *actually* passed over in a unit of time, since the motion is not uniform during that unit of time. Before a variable velocity can be measured, it is necessary to arrange that the motion *become uniform at a particular point or instant*, and that it *remain uniform for at least one succeeding unit of time.* The distance passed over during that unit of time then measures the velocity in the uniform motion, and also the velocity in the varied motion at the instant when it ceased to be varied and became uniform. Therefore, when the motion is varied, the velocity *at a particular instant* is the distance that *would be* passed over in one unit of time *if, beginning with that instant, the motion were to become uniform.*

The measurement of a velocity in varied motion at any particular point may be illustrated by allowing a ball that rolls down an inclined plane to leave the plane at that point and roll off upon a horizontal board. While the ball is on the inclined plane it has a variable velocity, because in this case its velocity is continually increasing; but as soon as it rolls off upon the horizontal board, it has approximately a constant velocity, and its motion is practically uniform. The actual velocity of the ball at the given point on the inclined plane may be determined by placing the end of the horizontal board at that particular point, and measuring the distance on this board that the ball passes over in one second of time. The distance passed over in one second on the horizontal board measures the velocity of the ball on this board, and therefore, also, its velocity on the inclined plane at the particular point at which it left the plane.

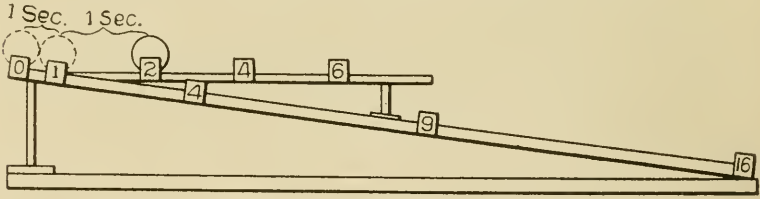

Fig. 8. Measurement of the velocity of a ball after rolling for 1 second on an inclined plane.

Thus, if the inclined plane is 16 feet in length and a little more than 1 foot in height, the ball will be observed to roll 1 foot in 1 second, 4 feet in 2 seconds, 9 feet in 3 seconds, and 16 feet in 4 seconds. Accordingly, during the first second it rolls 1 foot, during the second second 3 feet, during the third second 5 feet, and during the fourth second 7 feet; and, as it obviously passes over unequal distances in equal intervals of time, its motion is varied and its velocity is not constant. After it has rolled for 1 second on the inclined plane, Fig. 8, its velocity may be measured by placing the end of the horizontal board 1 foot from the top of the plane and observing the distance it rolls on this board during the succeeding second, which will be found to be 2 feet. The velocity of the ball on the horizontal board, and therefore also its velocity after rolling for 1 second on the inclined plane is thus found to be 2 feet per second.

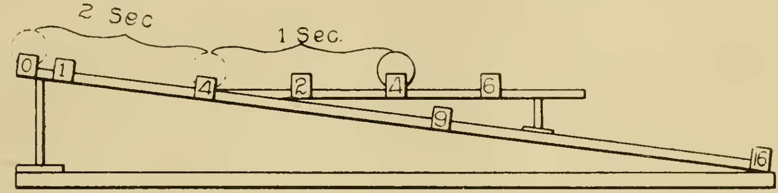

FIG. 9. Measurement of the velocity of a ball after rolling for 2 seconds on an inclined plane.

In the same way, Fig. 9, after the ball rolls for 2 seconds on the inclined plane, its velocity may be determined by observing the distance traversed in one second on the horizontal board, when this is placed 4 feet from the top of the plane. As it then passes over 4 feet in one second on the horizontal board, its velocity here, and therefore also its velocity after rolling for 2 seconds on the inclined plane, is 4 feet per second. Similarly, Fig. 10, after the ball rolls for 3 seconds on the inclined plane, its velocity is found to be 6 feet per second, and after it rolls for 4 seconds on the inclined plane, its velocity is 8 feet per second.

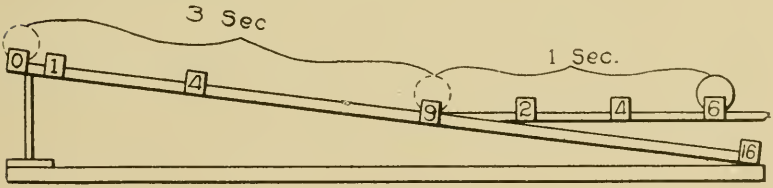

FIG. 10. Measurement of the velocity of a ball after rolling for 3 seconds on an inclined plane.

32. Acceleration. When the moving body passes over unequal distances in equal times, that is, when the velocity is changing in magnitude, or in direction, or in both, the motion is said to be *accelerated.* A change in the velocity, however, cannot take place instantaneously, and therefore a perfectly definite time is required for the velocity to change by a finite amount. If the velocity changes by the *same amount in equal successive intervals of time,* the motion is said to be *uniformly accelerated.* In this case, it is possible to divide the total change in the velocity by the time spent in producing this change, and so to obtain the quantity known as the *rate of change of the velocity,* or the *acceleration.* Accordingly, in uniformly accelerated motion, the *acceleration, or the rate of change of velocity,* is the ratio of the entire change of velocity to the entire number of units of time spent in producing this change, that is, it is the change of velocity that takes place in the unit of time. Since this change of velocity that takes place in one unit of time is the same as that that takes place in any other unit of time, the *acceleration is constant,* from which it follows that a *constant acceleration produces a uniformly accelerated motion.*

In order to determine the change of velocity that takes place during a certain time, we may subtract the velocity at the beginning of this time from the velocity at the end of the time. Thus if V_0* represents the velocity of the body when the observation begins, that is, the velocity at zero time, and V the velocity at the end of t units of time, then

$$V - V_0$$

represents the total change of the velocity in t units of time, and

$$\frac{V - V_0}{t}$$

represents the change of the velocity in one unit of time, or the acceleration. Therefore, if a represents the acceleration,

---

* According to the convention that will be adopted in these notes, v represents the *constant* velocity in *uniform* motion, and V the velocity at the end of the given time in *accelerated* motion.

$$a = \frac{V - V_0}{t} \qquad (1).$$

When the acceleration is not constant, this expression still holds true; but in this case t represents an extremely small interval of time, and the acceleration may be considered as remaining constant during this very short period.

33. The Measurement of Acceleration. The measurement of the constant acceleration in uniformly accelerated motion may be beautifully illustrated by the inclined plane and the horizontal board just described. If the ball starts from rest and rolls for one second on the inclined plane, its velocity at the end of that second will be 2 feet per second, as we have just seen. Its velocity at the beginning of the second being 0, and at the end of the second being 2 feet per second, the *change* of velocity that has taken place during that second is 2 feet per second, and this we have measured on the horizontal board. It was necessary, however, for the ball to roll for 1 second on the inclined plane in order to produce this change of velocity of 2 feet per second on the horizontal board. Accordingly, the *rate* of change of the velocity, that is, the *ratio of the change of velocity to the time spent in producing this change, or the acceleration,* is

$$\frac{\text{Change of velocity of 2 feet per second (on the horizontal board)}}{\text{1 second (on the inclined plane)}},$$

or the

$$\text{Acceleration} = \text{2 feet per second in a second.}$$

In the same way, if we allow the ball to roll for 2 seconds on the inclined plane, its velocity at the end of this time will be 4 feet per second, as we have seen above. The change of the velocity, as determined on the horizontal board, is thus found to be 4 feet per second, and it requires 2 seconds on the inclined plane to produce this change. The *acceleration,* therefore, or the ratio of the change of the velocity to the time spent in producing this change, is

$$\frac{\text{Change of velocity of 4 feet per second (on the horizontal board)}}{\text{2 seconds (on the inclined plane)}}$$

or,

$$\frac{\text{Change of velocity of 2 feet per second (on the horizontal board)}}{\text{1 second (on the inclined plane)}}$$

or,

Acceleration = 2 feet per second in a second.

In a similar way it may be shown that after the ball has rolled on the inclined plane for 3 seconds, and also for 4 seconds, its acceleration is still 2 feet per second in a second. It is thus seen that throughout the entire length of the inclined plane, the acceleration is *constant* and equal in magnitude to 2 feet per second in a second. Accordingly, a constant acceleration produces a uniformly accelerated motion.

It must be observed that this particular acceleration here taken for illustration is spoken of as 2 feet per second *in a second.* It must also be noted most carefully that these two seconds, which are always to be mentioned in speaking of an acceleration, are *not the same second.* The *first* second has to do with the change of the velocity—in our illustration as determined on the horizontal board—while the second second refers to *the time required for that change of velocity to take place,* or the time on the inclined plane.

Since the acceleration implies a change in the velocity of a body, that change may be such as to increase its motion, or to retard its motion, or to produce a change in the direction of its motion. Every motion, therefore, that is not a uniform motion in a straight line is an accelerated motion. Acceleration is positive or negative according as the velocity is increasing or decreasing. A negative acceleration produces a retardation of the motion.

34. The Unit of Acceleration. In the above expression for acceleration,

$$a = \frac{V - V_0}{t},$$

if both numerator and denominator are made equal to unity, the acceleration also becomes unity. The *unit acceleration,* therefore, is the acceleration of a body, *the velocity of which changes by one unit of velocity in one unit of time.* The unit of velocity being one centimeter per second, and the unit of time one second, the *unit of acceleration* is the acceleration of a body, *the velocity of which changes at the rate of one centimeter per second in a second.*

35. Distance Passed Over in Uniformly Accelerated Motion. When a body passes over a given distance in a certain time with accelerated motion, the velocity of the body changes from point to point. However varied may have been the motion during the time, it is evident that the *mean,* or *average,* of all the *varying* velocities is equivalent to the *constant* velocity with which the body *could* have passed over the same distance in the same time if the motion had been uniform and *not accelerated.* If v represents this *average constant* velocity, then, as in *actual* uniform motion (p. 24), the distance l passed over in time t is expressed by

$$l = v\,t.$$

If the motion is *uniformly* accelerated, the velocity increases or decreases by equal amounts in equal successive intervals of time, so that the *equivalent constant* velocity v is the average of the velocities V_0 and V, at the beginning and at the end of the time respectively, that is to say,

$$v = \frac{V_0 + V}{2}.$$

When this *value* of v replaces v in the above equation for the distance passed over in the *equivalent uniform motion,* we have, as the distance passed over in *uniformly accelerated motion,*

$$l = \frac{V_0 + V}{2} t. \qquad (2).$$

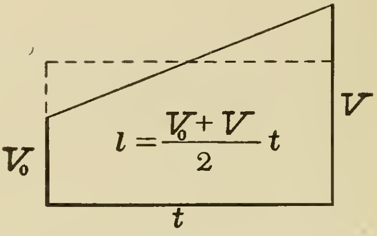

FIG. 11. In uniformly accelerated motion the distance passed over equals the average velocity multiplied by the time.

Just as the distance passed over by a body moving with *uniform* motion may be represented by the area of a rectangle (p. 25), so also may the distance described in *uniformly accelerated* motion be expressed by the area of a quadrilateral of the shape shown in Fig. 11. Here the velocities at the beginning and at the end of the time t are indicated by the vertical distances V_0 and V, respectively, while the *mean,* or *average,* velocity for the *entire* time is given by the mean height of the figure, which is

$$\frac{V_0 + V}{2}.$$

Since the mean, or average, height is now known, the area of the *equivalent rectangle* is obtained by multiplying this average height by the base. The distance l, therefore, passed over in *uniformly accelerated* motion is equal to the average velocity $\frac{V_0 + V}{2}$ multiplied by the time t, that is, it is equal to the average distance passed over in one second, multiplied by t, the number of seconds, or

$$l = \frac{V_0 + V}{2}$$

By referring to the last figure, it is to be observed that in uniformly accelerated motion the average velocity for the entire time is the same as the actual velocity at the end of the first half of the time.

36. The Laws of Uniformly Accelerated Motion. If V is eliminated between equations (1) and (2)

$$l = V_0 t + \frac{1}{2} a t^2; \qquad (3).$$

and if t is eliminated between (1) and (2),

$$l = \frac{V^2 - V_0^2}{2a}. \qquad (4).$$

These four equations represent the four great laws of uniformly accelerated motion, where the moving body has an initial velocity V_0.

If the body starts from rest, that is, if

$$V_0 = 0$$

these four equations reduce to the following:

$$a = \frac{V}{t}, \text{ or } V = a t, \qquad (5).$$

$$l = \frac{1}{2} V t, \qquad (6).$$

$$l = \frac{1}{2} a t^2, \qquad (7).$$

$$l = \frac{V^2}{2a}, \text{ or } V^2 = 2 a l. \qquad (8).$$

37. Falling Bodies. Under the influence of gravity, bodies falling freely experience an acceleration of 981 centimeters per second in a second. The acceleration, which is nearly uniform all over the earth, is constant for any one place and is known as the acceleration of gravity. Since falling bodies thus move with a *constant acceleration*, their motion as they fall is *uniformly accelerated*, and to them the foregoing laws of uniformly accelerated motion apply. In this case, however, the acceleration which is occasioned by the attraction of gravity is denoted by the letter g, where

$$g = 981 \text{ centimeters per second in a second.}$$

38. Bodies Falling with an Initial Velocity. If bodies fall with an initial velocity V_0, that is, if they are projected downward with this velocity, the motion, which is uniformly accelerated, is represented by equations (1)—(4), in which, however, g replaces a, as follows:

$$g = \frac{V - V_0}{t}, \text{ or } V = V_0 + g\,t, \qquad (9);$$

$$l = \frac{V_0 + V}{2}\,t, \qquad (10);$$

$$l = V_0\,t + \frac{1}{2}\,g\,t^2, \qquad (11);$$

$$l = \frac{V^2 - V_0^2}{2\,g}, \text{ or } V^2 = V_0^2 + 2\,g\,l. \qquad (12).$$

39. Bodies Falling Freely from Rest. If there is no initial velocity downward, that is, if the bodies fall freely from rest, then

$$V_0 = 0,$$

and the above four laws, (9)—12), become the laws of freely falling bodies, as follows:

$$g = \frac{V}{t}, \text{ or } V = g\,t, \qquad (13);$$

$$l = \frac{1}{2}\,V t, \qquad (14);$$

$$l = \frac{1}{2} g t^2, \qquad (15);$$

$$l = \frac{V^2}{2g}, \text{ or } V^2 = 2 g l, \qquad (16).$$

40. Projection Upward. The laws expressed in equations (9)—(12) are applied to the motion of bodies projected downward, where g is of the same sign as l, V_0, and V, all of which are positive in this direction. For bodies projected upward, l, V_0, and V are positive upward, while g, being still directed downward, is negative. The laws for projection upward, therefore, are the same as the above four laws for projection downward, in which, however, $-g$ replaces g, as follows:

$$V = V_0 - g t, \qquad (17);$$

$$l = \frac{V_0 + V}{2} t, \qquad (18),$$

$$l = V_0 t - \frac{1}{2} g t^2, \qquad (19);$$

$$l = \frac{V^2 - V_0^2}{-2g}, \text{ or } V^2 = V_0^2 - 2 g l. \qquad (20).$$

41. Problems in Projection Upward. These expressions for the motion of bodies projected upward make possible the solution of the following problems:

1. To find the time of ascent of a body when projected upward with an initial velocity of V_0.

The body will continue to ascend until the final velocity becomes zero and the body comes for an instant to rest, Fig. 12. In this case

$$V = 0,$$

and equation (17) becomes

$$V_0 = g t,$$

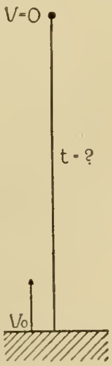

FIG. 12. Time of ascent.

or,

$$t = \frac{V_0}{g}.$$

This equation, however, is of the same form as (13), which expresses the time required by a body falling freely from rest to acquire the velocity V. The time, therefore, during which a body will continue to rise when projected upward with an initial velocity V_0 is the same as that required by a body when falling freely from rest to acquire a velocity equal to the initial velocity.

2. To find the height to which a body will ascend when projected upward with an initial velocity V_0.

In this case also, Fig. 13, the body will continue to ascend until the final velocity becomes zero, that is, until

$$V = 0.$$

Equation (20) then becomes

$$V_0^2 = 2\,g\,l,$$

or,

$$l = \frac{V_0}{2g}.$$

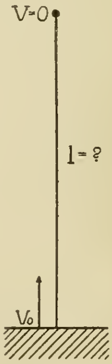

FIG. 13. Height of ascent.

But this equation is of the same form as (16), which expresses the height from which a body must fall freely from rest to acquire the velocity V. The height, therefore, to which a body will ascend when projected upward is the same as that from which the body must fall freely from rest to acquire a velocity equal to the initial velocity. The velocity of descent is thus shown to be the same as the initial velocity of ascent.

3. To find the time when a body, projected upward with an initial velocity V_0, will be at a given height.

The expression for the height in this case is given by equation (19),

$$l = V_0 t - \frac{1}{2} g t^2,$$

where l is the given height, Fig. 14. Since this is a quadratic in t, two positive, unequal values of t that will satisfy the equation may in general be found. This shows that there are two different times when a body will be at the given height, one during the ascent and the other during the descent.

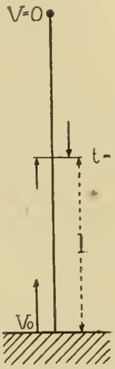

FIG. 14. Time at a given height.

CHAPTER IV.

VECTORS AND VECTOR QUANTITIES

42. Vectors. Any physical quantity that has direction as well as magnitude is called a *vector* quantity. Any quantity that possesses magnitude only, but not direction is a *scalar* quantity, or a scalar. Motion, velocity, acceleration, flow, etc., are examples of vectors, while mass, area, volume, time, energy, speed, etc., are scalars. Algebraic symbols, and, indeed, ordinary algebraic operations have to deal with scalar quantities only. Vector quantities, or vectors, may be represented by geometrical lines so drawn that each line is parallel in direction to the vector that it represents, and is also proportional in length to the magnitude of this vector. The minus sign before a vector indicates a negative vector. This, however, is the same as an equivalent positive vector, but turned in the opposite direction. Thus, vector BA, Fig. 15, is minus vector AB. Since magnitude and direction are the only quantities that specify a vector, a vector may be moved anywhere parallel to itself, provided only that its length and its direction remain unchanged.

FIG. 15. A Vector.

43. Addition of Two Vectors. If vectors are parallel to one another, they may be combined into a single vector by taking their algebraic sum. Thus, Fig. 16, vector *AB* plus vector *BC* equals vector *AC*. If these vectors represent motions, this indicates that motion from *A* to *B* plus motion from *B* to *C* is equivalent to motion from *A* to *C*. In the same way, vector *AC* minus vector *BC*, or, what is the same thing, vector *AC* plus vector *CB* equals vector *AB*. Again, if these vectors represent motions, this shows that motion from *A* to *C* plus motion from *C* to *B* is equivalent to motion from *A* to *B*.

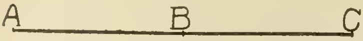

FIG. 16. Addition and subtraction of parallel vectors.

If vectors make an angle with one another, they cannot be united algebraically into a single vector, but may be combined or compounded only geometrically. In the case of two vectors, the line representing one is drawn from the end of the line representing the other, and the line now drawn from the starting point of the first vector to the end of the second is the "sum" of these two vectors. Thus, if vector *BC*, Fig. 17, is to be added to vector *AB*, a line equal and parallel to *BC*, Fig. 18, is drawn from the end of a line equal and parallel to *AB*. The final line *AC* that may now be drawn from the starting point *A* to the extremity *C* represents, in magnitude and direction, the sum of these two vectors. Thus:

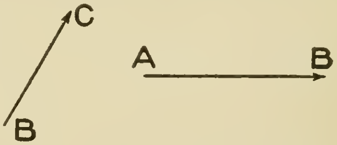

FIG. 17. Two vectors to be added.

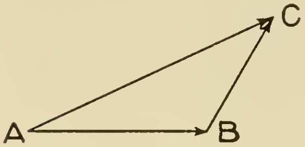

FIG. 18. Addition of vectors making an angle.

$$\text{Vector } AB + \text{Vector } BC = \text{Vector } AC.$$

This may seem at first to violate the proposition in geometry that the sum of two sides of a triangle is greater than the third. It must be remembered, however, that a wider meaning has been given to the sign $+$, and, therefore, an equally liberal interpretation must be extended to the sign $=$. The sign $=$ must, then, be read "equivalent to" instead of "equal to". The addition, therefore, of these two vectors must be understood to mean the following: a displacement from *A* to *B*, followed by a displacement from B to C, is equivalent to a displacement from A to C.

44. Addition of Several Vectors. If more than two vectors are to be compounded, this process must be continued until each vector in turn has been added to those that have preceded. Thus, if a third vector is to be compounded with these other two, a line, the length and direction of which represents this third vector, must be drawn from the end C of the second; a line representing a fourth vector must then be drawn from the extremity of the third; and so on. The line, that is finally drawn from the original starting point of the first vector to the extremity of the last, represents *in length and direction the magnitude of the sum* of the separate vectors. Thus if α, β, γ, ... Fig. 19, represent several vectors, AB, Fig. 20, may be drawn to represent α, BC may be drawn from B to represent β, CD may be drawn from C to represent γ, and so on. The line AG, which connects the starting point A of the first vector with the extremity of the last vector, is then the sum of the several vectors. The final sum of two or more vector quantities is called the *resultant* of the separate vectors, and the vectors themselves are called the *components*. The *resultant* of any number of vectors therefore, is *the single vector that can take the place of all the given vectors.*

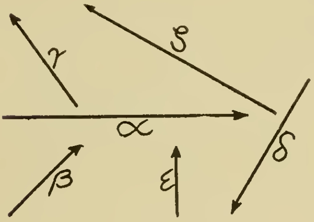

FIG. 19. Vectors.

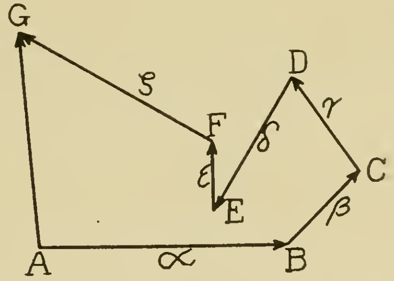

FIG. 20. Addition of several vectors.

It is to be observed that the component vectors, *taken in cyclical order*, form the sides of an open polygon, and that the line AG, representing the resultant of these vectors, is the line drawn from the point A that closes this polygon. The line closing the vector polygon must, however, be taken in a direction opposite to the cyclical order of the other sides of the polygon, that is, it must begin at the original starting point of the first vector. If the point G coincides with the point A, *the resultant is zero*, and the effect of all the vectors combined is the same as if no vectors whatever had operated. A system of vectors having a *zero resultant* produces the condition known as *equilibrium*. In forming the vector polygon, any vector may be considered as the first one, the component vectors may be taken in any order, and they may cross one another if necessary. Indeed, the polygon need not be plane.

45. The Parallelogram Law. The above means of combining vectors is equivalent to compounding them according to the parallelogram law. When the latter method is employed, and this is the one commonly used, lines representing the magnitudes and directions of the several component vectors are first drawn from a given point A, Fig. 21. Any vector as AB, for instance, is chosen as the first one, and any other vector as AD, or, what is the same thing, BC, is then added to it, Fig. 22, forming the sum AC, as we have just seen. Regarding then for the moment only these two vectors, AC, which is the sum of vectors AB and AD, is the diagonal of the parallelogram founded on vectors AB and AD as sides. Vector AC is, then, the resultant of vectors AB and AD, and these two vectors are regarded as the components of that resultant.

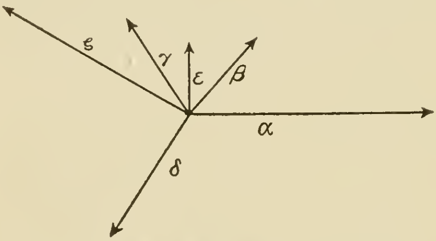

FIG. 21. The same vectors shown in Fig 19.

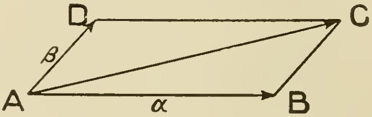

FIG. 22. Addition or composition of two vectors by the parallelogram law.

Since vector AC can now in every way *take the place* of vectors AB and AD, this single vector may be regarded as having actual existence, and may be compounded, also by the parallelogram law, with any other vector, as AF, Fig. 23. Accordingly, the resultant of vectors AC and AF is vector AE, and this is the diagonal of the parallelogram founded on vectors *AC* and *AF* as sides. Vector *AE*, which is the resultant of vectors *AC* and *AF*, is then, also the resultant of vectors *AB*, *AD*, and *AF*. In this way each resultant vector thus obtained may be compounded by the parallelogram law with another one of the component vectors until these components have all in turn been used. The final diagonal thus obtained represents, in magnitude and direction the resultant of all the component vectors.

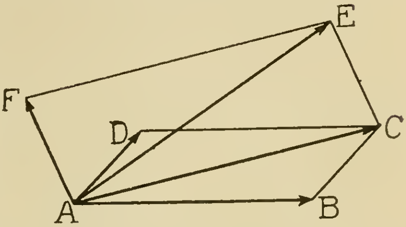

Fig. 23. Composition of three vectors by the parallelogram law.

46. The Resolution of a Vector. This method of compounding vectors has the further advantages of allowing *any vector* to be regarded as *the resultant of two component vectors, and of resolving this vector into these components.* That is, *any vector may be replaced by two components in any directions,* provided only that these components, when added together, make the original vector. In order that the vector may be resolved into its components, the *directions* of these components must be *specified,* but their *magnitudes* are determined by the *parallelogram law.* This is accomplished by regarding the given vector as the diagonal of a parallelogram, the sides of which are parallel to the required directions of the component vectors. The parallelogram is then completed by drawing lines from the extremity of this diagonal parallel to the given directions. The intercepts thereby produced upon these directions give the required magnitudes of the component vectors.

Thus, for instance, it may be required to resolve the vector *AC*, Fig. 24, into two components having the directions *AP* and *AQ*. From the extremity *C* of the given vector, lines *CB* and *CD* are drawn parallel to the given direction. The intercepts, *AB* and *AD*, upon these directions are the components required, for vectors *AB*

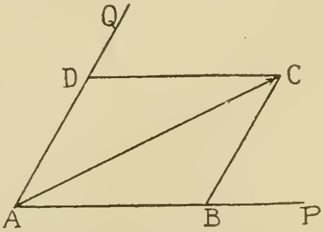

Fig. 24. Resolution of a vector into its components.

and *AD,* compounded according to the parallelogram law, are equivalent to vector *AC,* the resultant.

47. Resultant Motion—Composition of Two Uniform Motions. If two uniform motions are compounded according to the parallelogram law, the resultant is still a uniform motion, as may be seen with the aid of Fig. 25. Here *OA, OB, OC,* and *OA′, OB′, OC′* represent the distances described with uniform motion in 1, 2, and 3 units of time in the horizontal and in the inclined directions respectively, the equal distances *OA, AB, BC* and also the equal distances, *OA′*, A′B′, B′C′ being described in equal times. At the end of the first unit of time, if the horizontal motion only were present, the moving particle would be found at *A,* at the end of the second unit of time at *B*, and at the end of the third unit at *C,* and so on; while if the inclined motion were the only one to be considered, the particle at the end of the first unit of time would be found at *A′*, at the end of the second unit of time at *B′*, and at the end of the third unit at *C′*, and so on.

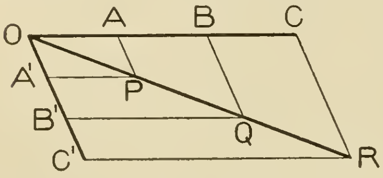

FIG. 25. The resultant of two uniform motions is a uniform motion.

The particle, however, is actuated by both motions simultaneously. Therefore, at the end of the first unit of time it is found by the parallelogram law to be neither at *A* nor at *A′*, but at P; at the end of the second unit, neither at *B* nor at *B′*, but at *Q;* at the end of the third unit, neither at *C* nor at *C′*, but at *R,* thus having passed over the distance *O P Q R.*

That this distance has been traversed with uniform motion may be seen by considering that, since the triangles *O A P, O B Q, O C R* are similar, their sides are proportional, so that

$$OP : OQ : OR = OA : OB : OC.$$

But

$$OA : OB : OC = 1 : 2 : 3,$$

and therefore

$$OP : OQ : OR = 1 : 2 : 3.$$

In other words, the distances passed over are proportional to the times spent in traversing these distances, and so (p. 24, eq. 2), the motion from O to R is uniform. Accordingly, two uniform motions compounded by the parallelogram law produce a resultant uniform motion.

48. Composition of Two Uniformly Accelerated Motions. In a similar manner it may be proved that, if two uniformly accelerated motions are compounded by the parallelogram law, the resultant motion is still uniformly accelerated. Let OA, OB, OC and OA', OB', OC', Fig. 26, be the distances described with uniformly accelerated motion in 1, 2, and 3 units of time in the horizontal and in the inclined directions respectively, where

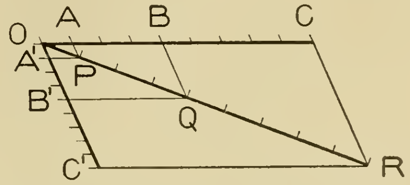

FIG. 26. The resultant of two uniformly accelerated motions is a uniformly accelerated motion.

$$OA : OB : OC = OA' : OB' : OC' = 1 : 4 : 9.$$

Because the triangles OAP, OBQ, OCR are similar, we see that

$$OP : OQ : OR = OA : OB : OC;$$

and, as

$$OA : OB : OC = 1 : 4 : 9,$$

it follows that

$$OP : OQ : OR = 1 : 4 : 9.$$

Since the distances between O and R are thus proportional to the squares of the times spent in describing them, the motion from O to R must be uniformly accelerated (p. 33, eq. 7). Therefore, two uniformly accelerated motions compounded according to the parallelogram law yield a resultant motion which is likewise uniformly accelerated.

49. Composition of a Uniform Motion and Uniformly Accelerated Motion. If uniform motion in one direction is compounded with uniformly accelerated motion in another, the resultant motion does not take place along a straight line, but along a curve known as a parabola. When a bullet is fired horizontally from a gun, its inertia tends to make it move with uniform motion in the horizontal direction, while as soon as it leaves the mouth of the gun it becomes a freely falling body, and so tends to move with uniformly accelerated motion vertically downward. The actual path which it follows in its flight is the resultant of these two motions.

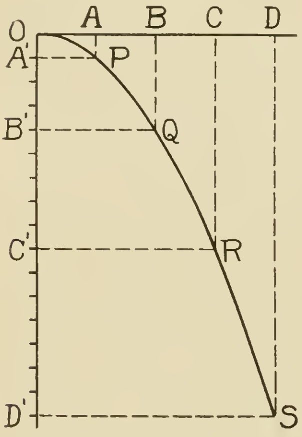

FIG. 27. A uniform motion compounded with a uniformly accelerated motion at right angles.

If the horizontal motion only were present, the bullet would describe the horizontal distance.

$$x = v\,t \qquad (1)$$

with uniform motion (p. 24, eq. 2), and at the end of 1 second would be found at *A*, Fig. 27, at the end of 2 seconds at *B*, at the end of 3 seconds at *C*, and so on. Similarly if the motion downward were the only one existing, it would describe the vertical distance

$$y = \frac{1}{2}\,g\,t^2 \qquad (2)$$

with uniformly accelerated motion (p. 35, eq. 15), and at the end of 1 second would be found at *A′*, at the end of 2 seconds at *B′*, and so on. Since it is actuated by both of these motions at the same time, it is found by the parallelogram law to be, at the end of the first second neither at *A* nor at *A′*, but at *P;* at the end of the second second neither at *B* nor at *B′*, but at *Q;* at the end of the third second neither at *C* nor at *C′* but at *R,* and so on, thus having traversed the curved path *O P Q R S*.

The distances described with uniform motion in the horizontal direction—the abscissae—are proportional to the times 1, 2, 3, 4. . . ., while the distances described with uniformly accelerated motion in the vertical direction—the ordinates—are proportional to the squares of the times, that is, to 1^2, 2^2, 3^2, 4^2 . . . ., or 1, 4, 9, 16 . . . . ; and therefore the ordinates are proportional to the squares of the abscissae. The curve in which this relation holds true is a parabola, the axis of which is vertical. Accordingly, if the resistance of the air were neglected, a bullet fired horizontally from a gun, or a jet of water held horizontally, would describe a parabolic path. The curve, indeed, would remain a parabola, although of different form, if the bullet or the jet of water were to make any angle with the horizontal.

If t is eliminated between equations (1) and (2), (p. 44), we have the expression

$$x^2 = \frac{2\,v^2}{g}\,y,$$

which shows the relation between the horizontal and the vertical distances for every point of the parabolic path. By assigning different values to x, the corresponding values of y may be calculated, and so the curve may be plotted point by point.

50. The Inclined Plane. When a body falls freely under the action of gravity, the acceleration influencing its motion acts in the direction of the motion itself, and therefore has the value of 981 centimeters per second in a second. When, however, a body or particle moves down an inclined plane, the vertical acceleration acts at an angle with the face of the plane, and therefore at an angle with the direction of motion. Accordingly, the change in the magnitude of the velocity is due only to the component of the acceleration which acts in the direction of motion, that is, parallel to the face of the plane. Thus, if ABC, Fig. 28, represents an inclined plane, a particle placed upon it would have a total acceleration of *DE*, or *g*, acting vertically downward if this particle were perfectly free. Inasmuch, however, as it is not free, this acceleration may be resolved into two components, one *DF* parallel to the face of the plane, and the other *DG* perpendicular to this face. The former of these components *DF*, acting in the direction of motion, causes the change in the velocity of the body as it moves down the plane. The latter component *DG*, acting perpendicularly to the face of the plane, has no part or component in the direction of motion, and so cannot affect the motion of the particle. If the angle of the plane is denoted by ϕ, the angle *DEF* is also ϕ, since its sides are perpendicular to those of the plane, and

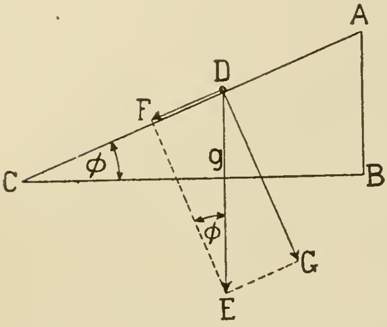

FIG. 28. Acceleration on the inclined plane.

$$\frac{DF}{DE} = \frac{DF}{g} = \sin\phi,$$

or,

$$DF = g \sin\phi.$$

The body, therefore, moves down the inclined plane, not with the entire acceleration *g*, but with the fractional part of this acceleration denoted by sin ϕ, that is, with the acceleration $g \sin\phi$.

The four expressions, (13)—(16) (p. 35), for bodies falling freely from rest assume, then, the following forms when they refer to motion down an inclined plane:

$$g \sin\phi = \frac{V}{t}, \text{ or } V = g\,t \sin\phi, \qquad (21);$$

$$l = \frac{1}{2}\,V\,t, \qquad (22);$$

$$l = \frac{1}{2}\,g\,t^2 \sin\phi, \qquad (23);$$

$$l = \frac{V^2}{2\,g\,\sin\phi}, \text{ or } V^2 = 2\,g\,l\,\sin\phi. \qquad (24).$$

In these expressions l represents the distance through which the body moves along the face of the plane.

51. To Find the Velocity with which a Body, Moving Down an Inclined Plane, Reaches the Foot of the Plane.

If h, Fig. 28, denotes the height AB of the inclined plane, and l the length AC of this plane, then

$$\sin\phi = \frac{h}{l}.$$

Equation (24), which expresses the velocity of the body as this reaches the foot of the inclined plane, then becomes

$$V^2 = 2\,g\,l\,\frac{h}{l},$$

or,

$$V^2 = 2\,g\,h.$$

Since this equation is independent both of l and of ϕ, it is evident that the velocity which a body acquires in moving down an inclined plane is independent both of the length of the plane and of the angle of the plane. Moreover, from equation (16), (p. 35), we learn that the velocity acquired by a body falling freely through a distance h, equal to the height of the plane, is

$$V^2 = 2\,g\,h.$$

Inasmuch as these two expressions are the same, it is seen that the velocity acquired by a body moving down the face of an inclined plane is the same as that which would be acquired by the body in falling freely through a distance equal to the height of the plane. Therefore, the velocity with which a body reaches the foot of an inclined plane is independent of the length of the plane and of the angle of the plane, and is the same in amount as that acquired by a body in falling through a distance equal to the vertical height of the plane.

CHAPTER V.

MOTION OF ROTATION.

52. Rotation. Another type of motion, in addition to the linear motions we have heretofore considered, is found in the *rotation* of a particle or of a rigid body. If a particle rotates, it must turn about a straight line, or axis, not passing through the particle itself; and if a rigid body rotates, it likewise must turn about an axis which may or may not pass through the body. When a body rotates, each particle of which it is composed moves along the circumference of a circle, the center of which lies upon the axis of rotation.

53. The Unit Angle, or the Radian. In motions of rotation, the angles through which the rotating bodies turn are seldom measured in degrees, but are expressed in terms of the *unit angle,* or the *radian.* In circular measure, the *unit angle, or the radian, is the angle subtended at the center of a circle by an arc equal in length to the radius.* In other words, if the radius of a circle were bent along the circumference, the arc so formed would subtend the unit angle. Therefore, in the unit angle, the arc and the radius are of equal length. From this it follows, Fig. 29, that, if a rotating body turns through the unit angle, a particle 1 centimeter from the axis describes an arc 1 centimeter long, and a particle r centimeters from the axis describes an arc r centimeters long. Moreover, Fig. 30, if θ represents the number of radians through which a body rotates, θ is also the number of centimeters in the length of the arc 1 centimeter from the axis, and $r\theta$ is the number of centimeters in the length of the arc r centimeters from the axis. In general, if a represents the length of an arc,

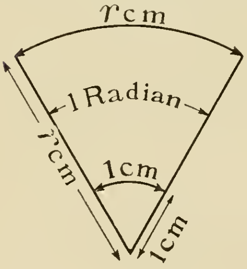

FIG. 29. One radian.

$$a = r\,\theta,$$

where r is the length of the radius, and θ is the number of radians through which the body rotates.

54. The Value of the Radian in Degrees. From very early attempts made to express the length of an entire circumference, it was found that the circumference of a circle bears to its diameter the incommensurable ratio of π, where

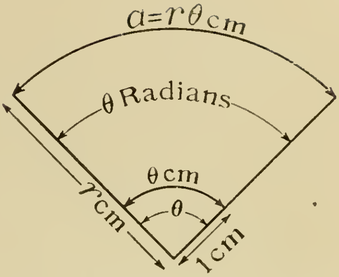

FIG. 30. Measurement of an arc in circular measure.

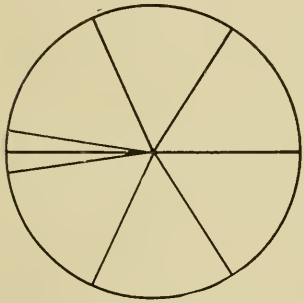

FIG. 31. The entire circumference of a circle subtends an angle of 2π radians.

$$\pi = 3.141592 \ldots,$$

so that

$$c = \pi d,$$

or

$$c = 2\pi r,$$

where c is the circumference, d the diameter, and r the radius of the circle.

By a comparison of this expression for the length of the circumference with the above general equation for the length of an arc, we find, Fig. 31, that the angle at the center of a circle subtended by the entire circumference is

$$2\pi \text{ radians.}$$

Inasmuch as this angle in our ordinary degree-measurement is indicated by 360°, we see that

$$2\pi \text{ radians} = 360°,$$

and that

$$1 \text{ radian} = 57^\circ \; 17' \; 44.9''.$$

55. Angular Velocity. When a rigid body rotates, or turns upon an axis—a wheel, for instance,—it is evident that particles remote from the axis move through a greater linear distance than do those which lie near the axis about which the body rotates. In other words, the speed, or linear velocity, of a particle in a rotating body becomes greater the farther removed this particle may be from the axis of rotation. Since all parts of a rigid body turn through the same angle, it is often more convenient to speak of the *angular velocity of the entire body* than of the *linear velocity of any one of its particles.* By the angular velocity of a body which rotates uniformly, we mean the ratio of the entire angle through which the body turns to the entire number of units of time spent in turning through this angle; that is, *the angular velocity is the angle through which the body turns in one unit of time.* Therefore, the angular velocity of one particle of a rotating body is the same as that of any other particle, and of the body itself, since these all turn through the same angle in the same time.

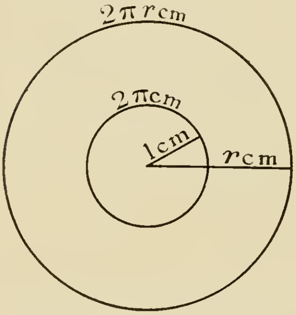

FIG. 32. The length of a circumference.

If a rotating body makes one complete revolution, that is, if it turns through 2π radians, Fig. 32, a particle 1 centimeter from the axis describes a circumference.

$$2\,\pi \text{ centimeters}$$

in length, and a particle r centimeters from the axis describes a circumference

$$2\,\pi\, r \text{ centimeters}$$

in length. If T is the time of one revolution of the rotating body, that is, the time of turning through 2π radians, the *linear velocity* v of a particle 1 centimeter from the axis is

$$v = \frac{2\,\pi}{T} \text{ centimeters per second,}$$

and the *linear velocity* v of a particle r centimeters from the axis is

$$v = \frac{2\,\pi\,r}{T} \text{ centimeters per second.}$$

At the same time, the *angular velocity* ω of the entire *rotating body* is the ratio of the *angle* through which the body turns during one revolution, or $2\,\pi$ radians, to the number of units of time T spent during this revolution, or

$$\omega = \frac{2\,\pi}{T} \text{ radians per second.}$$

In other words, *the angular velocity of a rotating body is the number of radians through which the body turns in one unit of time.*

56. Curvilinear Motion. When a particle moves along a continuous curve, it is moving for an instant at any point of the curve in the direction of the tangent at that point. Since the tangent at an adjacent point lies in a different direction, the particle, which then moves in the direction of this new tangent, is moving in a direction different from that at the previous point. The direction of the *motion* of the particle having thus changed, the direction of the *velocity has likewise changed.* But *any change in the velocity,* whether this be a change in direction only, or in magnitude, or in both, implies an *acceleration;* and therefore, since the *direction* of the velocity is changing from point to point, the motion of a particle along a curved path is *accelerated.*

Thus, Fig. 33, if a particle moves with uniform motion from A to B, no change occurs either in the magnitude or in the direction of the velocity, hence for the first part of the path there is no acceleration. While the particle moves from *B* to *C*, however, the velocity changes in direction from *AB* to *CD*. Since a change occurs in the velocity over this part of the path, an acceleration must exist, one component of which, at least, must lie at right angles to the direction of motion. At *C* the velocity ceases to change, the acceleration then becomes zero, and from *C* to *D* the motion is once more uniform.

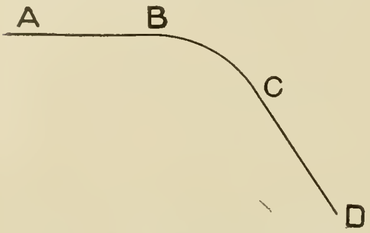

Fig. 33. Acceleration in curvilinear motion.

The acceleration which is thus present in all curvilinear motion illustrates the statement previously made (p. 30), that "every motion which is not a uniform motion in a straight line is an accelerated motion.

57. Constant Central Acceleration in Uniform Circular Motion. When a particle moves uniformly, that is, with constant speed, in the circumference of a circle, the velocity with which it moves coincides in direction at every point with the direction of the tangent at that point. Since the direction of the tangent changes *continuously*, the direction of the velocity changes likewise continuously, and the motion is therefore accelerated. Moreover, as the speed of the particle remains constant in its circular path, it is evident that no component

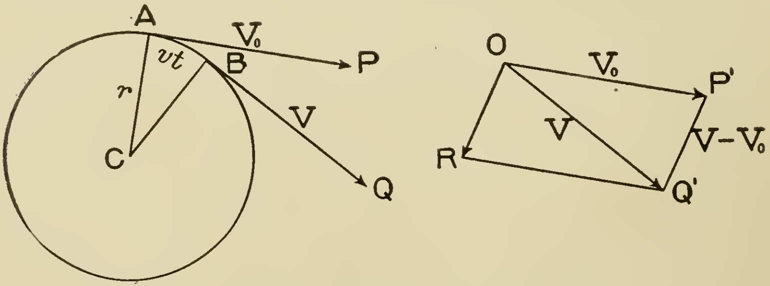

Fig. 34. Constant central acceleration in uniform circular motion.

of this acceleration lies in the direction of motion, from which we infer that the entire acceleration is at right angles to this direction, that is, it is directed towards the center of the circle. That this acceleration is not only constant in amount, but also directed at every point towards the center of the circle, is proved in the following demonstration:

A body, moving uniformly with the constant speed v in the circumference of the circle, Fig. 34, passes over the distance from A to B, where B may be taken so near to A that the arc AB coincides with the chord AB. That is to say, the distance AB may be taken as one of the infinite number of straight lines of which the circumference of the circle is composed. If this distance is described in t units of time, then

$$AB = vt, \qquad \text{(p. 24)}$$

since v is the speed, or the distance along the arc which is passed over in one unit of time. At A the velocity of the body is represented in magnitude and direction by the line AP, or V_0, and at B by the line BQ, or V. *V_0 and V differ from each other only in direction,* and both are *numerically* equal to the constant speed v.

Inasmuch as velocities are vector quantities, we may determine the change in the velocity that takes place between A and B by choosing some point, as O, and by laying off the lines OP' and OQ', proportional in length and parallel in direction to V_0 and V respectively. It is now evident that *the velocity V is the final effect of the initial velocity V_0 and of the change which the latter undergoes;* and so the vector V must be the resultant of two vectors, of which one is the vector V_0, while the other is the vector which, when added to the Vector V_0, will produce the vector V. By completing the parallelogram, in which one side is V_0 and the diagonal is V, we see that the vector OR, or its equal $P'Q'$, is the vector that must be added to the vector V_0 to produce vector V, that is,

$$\text{vector } V_0 + \text{vector } OR = \text{vector } V,$$

from which

$$\text{vector } OR = \text{vector } V - \text{vector } V_0,$$

or

$$\text{vector } O\,R = \text{vector } V - V_0.$$

From this we see that $V-V_0$, which is the change in the velocity that takes place between A and B, is represented in magnitude and direction by the vector $O\,R$.

We have now to determine the magnitude as well as the direction of this vector, and this may be accomplished as follows: Since the radius and the tangent are perpendicular to each other, the triangles $P'OQ'$ and $A\,C\,B$ are similar and their sides are proportional. Therefore,

$$\frac{P'\,Q'}{A\,B} = \frac{O\,P'}{C\,A},$$

or

$$\frac{V - V_0}{v\,t} = \frac{V_0}{r},$$

from which

$$\frac{V - V_0}{t} = \frac{v\,V_0}{r}.$$

But the left hand member of this equation is *the rate of change of the velocity, or the acceleration a,* that is

$$\frac{V - V_0}{t} = a, \qquad \text{(p. 29)}$$

while the right hand side contains V_0, which is numerically equal to v. By replacing these values in the above equation, we find that

$$a = \frac{v^2}{r}.$$

Here v^2 and r are constant quantities, and hence a is likewise *constant in magnitude.*

Moreover, the *direction* of this acceleration is given by the direction of the change of the velocity, and this is indicated by the direction of the vector $O\,R$. When the point B, however, is *as near as we please* to the point A, the angle ACB *is as small as we please*, and the angle $P'\,O\,Q'$, or its equal $O\,Q'\,R$, is also *as small as we please.* Therefore, the angle which the vector $O\,R$ makes with the vector $O\,P'$, or V_0, is then *as near as we please* to a right angle; and *in the limit*, when the point B is taken *infinitely* near to the point A, the vector $O\,R$ is then strictly at right angles to the vector OP', or V_0. In other words, the acceleration, since it is at right angles to the tangent AP, or V_0, is directed towards the center of the circle.

From this we see that when a particle revolves uniformly, that is, with constant speed, in the circumference of a circle, the motion is accelerated. This acceleration a is always *constant in amount*, and equal to the square of the speed divided by the radius, or

$$a = \frac{v^2}{r},$$

and is always directed towards the center of the circle.

The *magnitude* of this central acceleration may be expressed, as above, in terms of the speed v of the particle, that is, in terms of the actual distance along the arc which this particle travels in the unit of time; or, it may be expressed, as follows, in terms of the *angular* velocity ω, (p. 51). If T represents the time of one complete revolution

$$v = \frac{2\,\pi\,r}{T}, \qquad \text{(p. 51)}$$

so that the magnitude of the acceleration which acts towards the center, and which is called the centripetal acceleration, may then be written.

$$a = \frac{v^2}{r} = \frac{4\,\pi^2\,r^2}{T^2\,r} = \frac{4\,\pi^2\,r}{T^2},$$

But

$$\frac{4\,\pi^2}{T^2} = \omega^2, \qquad \text{(p. 51)}$$

and therefore

$$a = \omega^2\,r.$$

CHAPTER VI.

HARMONIC MOTION.

58. Harmonic Motion. If a particle moves uniformly in the circumference of a circle, and is observed from a distant point in the plane in which the circle lies, this particle appears no longer to move in a circular path, but seems to travel back and forth in a straight line across the direction of vision. Thus, if the eye is placed at the bottom of the page in the direction $Y\,Y'$, Fig. 35, the particle P, which moves uniformly around the circumference of the circle $X\,Y\,X'$, apparently moves back and forth along the diameter $X\,X'$. At X it is actually moving directly away from the observer; but, as this motion cannot be detected by the eye, the particle seems for an instant to stand still at this point. At Y, however, it is moving at right angles to the line of vision, and therefore appears to be at O, moving to the left.

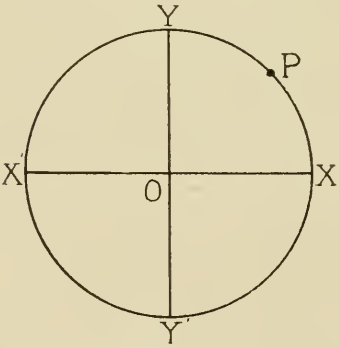

Fig. 35. Production of harmonic motion.

For points intermediate between X and Y, the motion of P in the circle may be resolved into two components, one of which lies in the line of vision, while the other is perpendicular to this direction. Since the eye cannot detect a motion in the line of vision, it is only the latter of these two components that can be observed. Moreover, this component increases in magnitude as point Y is approached, so that the particle apparently has an increasing velocity from X to O. For the same reason, while the particle really moves uniformly from Y to X' in its circular path, it appears to move with continually decreasing velocity along the straight line from O to X' and at X', where it is moving directly towards the observer, it apparently comes again to rest. During the remainder of the revolution from X' through Y' to X, the particle increases in velocity as it apparently goes from X' to O, then again decreases in velocity from O to X, and at this point it appears once more to be stationary. The particle has actually made one complete revolution in the circumference of the circle moving uniformly; but to the eye placed at the botton of the page it has *apparently* moved back and forth along the diameter XX', coming to rest at X and X', and moving with greatest velocity at O.

Back-and-forth, or oscillatory, motion, such as this, is of frequent occurrence in many branches of physics, and is known as simple harmonic motion. The pendulum of the common clock, for example, except that it moves in an arc instead of a straight line, executes simple harmonic motion as it vibrates to and fro; the particles of an elastic body move with harmonic motion as the body regains its original form after distortion; the strings of a musical instrument vibrate with harmonic motion, as do also the particles of air that transmit these vibrations to the ear; the particles of ether are thought likewise to vibrate with harmonic motion as heat and light come to us by radiation from the sun. The frequent occurrence of harmonic motion, and the aid it furnishes in explaining many natural phenomena, render it one of the principal motions with which the science of physics has to deal.

59. Terms and Definitions Used in Harmonic Motion. In the motion described above, the movement of P in its circular path is everywhere *projected* upon the diameter XX'. *Harmonic motion,* therefore, *is the projection upon a diameter of motion that takes place uniformly in the circumference of a circle.* It may also be defined *as the motion of the foot of the perpendicular let fall upon the diameter of a circle from a particle that moves uniformly in the circumference.* Thus, if the particle, P, Fig. 36, moves uniformly in the circumference of the circle XYX' and XX' is any diameter, p, or the foot of the perpendicular let fall from P upon the diameter, moves with harmonic motion. While P, Fig. 37, moves from X to A, p moves from X to a. While P moves from A to B, p moves from a to b, and so on. While P makes one complete revolution around the circle, p makes one complete excursion forward and back across the diameter. As long as P continues to revolve in the circle, p continues to execute harmonic motion, making one complete vibration forward and back for every revolution of P.

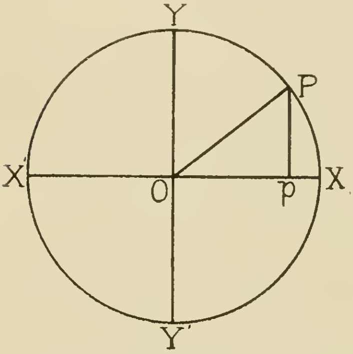

FIG. 36. Harmonic motion is the projection of circular motion.

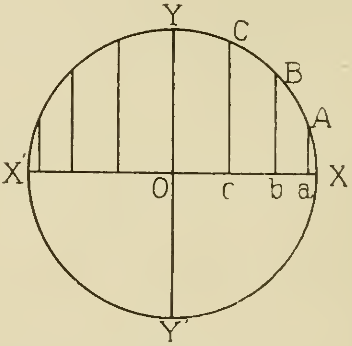

FIG. 37. Harmonic motion and circular motion.

The circle in which P revolves uniformly is called the *auxiliary circle, or the circle of reference,* and the *path* in which the harmonic motion takes place is always *a diameter of the auxiliary circle.*

The *period* of a particle executing harmonic motion is the time of one complete oscillation forward and back along the path, for instance, the time required for p to move from X to X' and back again to X. If a point other than the extremity of the path is taken as the starting point, *the period is the interval* *that elapses between two successive passages of the particle over the same point, moving in the same direction.* Therefore, the period of a particle vibrating with harmonic motion is the same as the time required by the particle which moves in the auxiliary circle to make one complete revolution.

The *frequency* is the number of complete vibrations made by the moving particle in the unit of time. Obviously this number increases as the time of one vibration decreases, and so the frequency is the reciprocal of the period. Accordingly, if n denotes the frequency and T the period, the relation existing between n and T is given by the equation,

$$n = \frac{1}{T}.$$

Motion to the *right* along the path is regarded as *positive,* and *motion* to the *left* is considered as *negative. Distances* to the *right of the center* of the path are *positive,* while *distances* to the *left* of this point are *negative.* The center of the path is called *the position of equilibrium.*

The distance of the moving particle at any instant from the position of equilibrium is known as the *displacement.* The *displacement is positive* when the moving particle is to the *right* of the position of equilibrium, and *negative* when it is to the *left* of this point.

The *amplitude* is the distance from the position of equilibrium to the extremity of the path in either direction, and is always equal to the radius of the circle of reference. Accordingly, the amplitude is the greatest possible displacement. The *maximum elongation* of the particle is the position the particle occupies when it is at either extremity of its path.

The phase is the fractional part of the period that has elapsed since the particle was last at the position of positive maximum elongation. The phase, therefore, is always *a proper fraction* and always *an abstract number.* For example, when the particle is at X, the phase is zero, since it is at the point from which phase is reckoned. When the particle is at O, moving to the left, the phase is $\frac{1}{4}$, inasmuch as one quarter of the period has elapsed since the particle left the point X. In the same way when the particle is at X', the phase is $\frac{1}{2}$; and when it is at O, moving to the right, the phase is $\frac{3}{4}$. When the particle reaches the point X, the phase, which has been approaching the value 1, now drops again to zero, and its former values are repeated during the following vibration of the particle and during all succeeding vibrations.

Omit theory

60. The Equation of Harmonic Motion. The back-and-forth movement of the particle p, Fig. 38, as this executes harmonic motion, may be represented by an equation in which the displacement of p is expressed as a function of the position of the particle P in the auxiliary circle. The position of P depends upon the angle θ, which is known as the phase angle, and which is the angle described by the radius vector since P left the line OX, or the X axis. That is, during the time in which the particle in the auxiliary circle has moved from X to P, the radius vector has moved from OX to OP, thus sweeping out the angle θ radians.

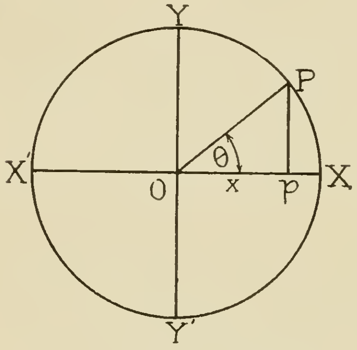

FIG. 38. Displacement of the particle in harmonic motion.

In this time the particle p, which is executing harmonic motion, has moved from X to p, and has at this instant the displacement Op, or x. Accordingly, the position of p in its path is connected with the position of P in the circle by the relation

$$\frac{x}{OP} = \cos\theta,$$

or

$$x = OP\cos\theta.$$

But OP, or its equivalent OX, is the radius of the auxiliary circle, and this, in turn, is the amplitude of the harmonic motion. Accordingly, if

$$A = \text{amplitude},$$

then

$$x = A \cos \theta.$$

Since this equation expresses the relation between the position of p in its path and the corresponding angle through which P has revolved in the auxiliary circle, the displacement of the particle executing harmonic motion is hereby defined for every value of θ. Since this is so, this expression is known as *the equation of harmonic motion*. As P continues to revolve, the angle θ increases uniformly; and as this angle assumes different values, the displacement of p may be seen to change as follows:

$$\text{When } \theta = 0 \text{ radians, } x = A,$$

$$\theta = \frac{\pi}{2} \text{ radians, } x = 0,$$

$$\theta = \pi \text{ radians, } x = -A,$$

$$\theta = \frac{3\pi}{2} \text{ radians, } x = 0,$$

$$\theta = 2\pi \text{ radians, } x = A.$$

This shows that the motion of the particle executing harmonic motion is from one extremity of its path through the center to the opposite extremity and back again, and that this motion is repeated as often as these values of θ recur.

61. The Phase. If t denotes the time of describing θ radians, Fig. 39, and if T is the period, that is, the time of describing one entire circumference, or 2π radians, t bears the same relation to T that θ bears to 2π, so that

$$t : T = \theta : 2\pi,$$

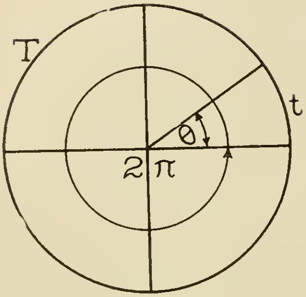

FIG. 39. Value of the angle θ.

or,

$$\theta = \frac{2\pi t}{T}.$$

The equation of harmonic motion then becomes

$$x = A \cos \frac{2\pi t}{T}.$$

Here t is the variable quantity upon which the value of x depends, and represents the time that has elapsed since the particle executing harmonic motion was last at the point X. Therefore, the fraction $\frac{t}{T}$ represents the *phase*, and this assumes different values as the particle occupies different points of the path. When the values assigned to the phase $\frac{t}{T}$ correspond to those given above to the phase-angle θ, the same series of displacements will be obtained. That is,

When $\frac{t}{T} = 0$, or zero phase, $x = A$,

$\frac{t}{T} = \frac{1}{4}$, or $\frac{1}{4}$ phase, $x = 0$,

$\frac{t}{T} = \frac{1}{2}$, or $\frac{1}{2}$ phase, $x = -A$,

$\frac{t}{T} = \frac{3}{4}$, or $\frac{3}{4}$ phase, $x = 0$,

$\frac{t}{T} = 1$, or zero phase, $x = A$.

62. The Velocity of a Particle Moving with Harmonic Motion. If perpendiculars are dropped at equal intervals of time from the particle P, Fig. 40, as this revolves in the auxiliary circle, the intercepts of these perpendiculars upon the diameter represent the distances described in equal times by the particle moving with harmonic motion. These intercepts are found to be of unequal length, being greater near the central portion of the diameter than towards the extremities. Since these *unequal distances* are described in *equal intervals of time,* the *velocity* of a particle moving with harmonic motion *is not constant,* but varies with the position of the particle in its path.

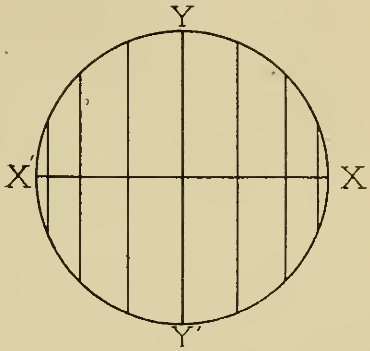

FIG. 40. Variable velocity in harmonic motion.

To determine the value of this variable velocity at any instant, we must remember that the *motion* of p in its path is the projection upon a diameter of the *motion* of P in the circle; and from this we conclude that the *velocity* of p is likewise the projection upon the same diameter of the *velocity* of P. In other words, *the velocity of p is the component of the velocity of P that is parallel to the diameter along which p moves.* But the velocity of P at every point is in the direction of the tangent to the circle at that point (p. 51). Accordingly, the line v, drawn tangent to the circle at P, Fig. 41, represents the velocity of P at this particular point, inasmuch as the direction of this line coincides with the direction of the velocity, and the length of this line is proportional to the magnitude of the velocity, that is, to the magnitude of the constant speed v, with which P moves in the circumference of the circle.

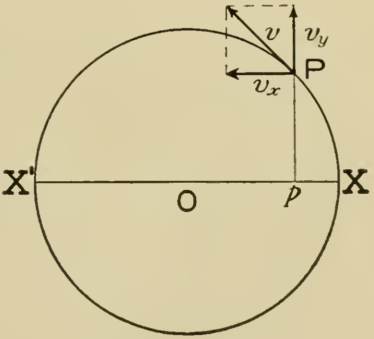

FIG. 41. Determination of the velocity in harmonic motion.

Since the horizontal diameter has been chosen as the path of the harmonic motion, it is necessary to resolve the velocity v into two components, one v_x in the horizontal direction, and the other v_y in the vertical direction. However, these components may vary in magnitude or direction as P revolves in the circle, the horizontal component v_x represents at any instant the velocity of p in its motion along the horizontal diameter, while the vertical component v_y expresses the velocity at the same instant of a particle moving along the vertical diameter.

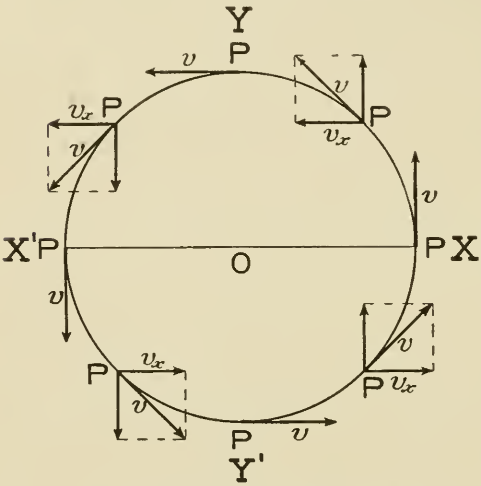

FIG. 42. Determination of the velocity at different points of the path.

By resolving the velocity of P into its components at different points on the circle of reference, we may determine the variations which occur in the velocity of p at corresponding points of the path. Thus, at the position X, Fig. 42, the velocity of P is directed vertically upward, so that at this point the horizontal component is zero; hence the velocity of p at the extremity of its path is also zero, and p is for an instant here at rest. As P moves in its circular path from X to Y, the horizontal component, which at first is exceedingly small, becomes continually greater, thus indicating that p moves towards O with an increasing velocity. At Y, the velocity of P is horizontal, and so coincides in value with its horizontal component, the vertical component here being zero. Accordingly, the velocity of p at O is for an instant constant, and equal in magnitude to the uniform speed of P in the auxiliary circle. As P moves from Y to X', the length of the horizontal component decreases, and becomes zero at X', since v is here directed vertically downward.

Therefore, the velocity of *p* decreases from *O* to *X′* and becomes zero at this extremity of its path, when *p* is again at rest. As *P* continues its motion for the remaining half-revolution, that is, from *X′* through *Y′* to *X*, these values of the horizontal component are repeated in the reverse order and in the opposite direction.

These variations of the velocity of *p*, which are repeated for each complete revolution of *P* in the circle of reference, are represented for a single period by the velocity vectors in the accompanying diagram, Fig. 43. It is to be observed that the velocity of *p* at any point is always a fractional part of the velocity of *P*, this fraction varying in value from zero at the extremities of the path to unity at the center. Inasmuch, however, as the velocities are directed to the left, or in the negative direction, for the first half of the period, and to the right, or in the positive direction, for the second half, it is necessary always to use the minus sign when the velocity for the first half of the period is indicated.

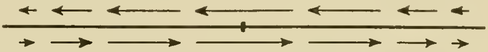

FIG. 43. Variation of the velocity for an entire period.

63. The Value of the Velocity at any Point. With the aid of Fig. 44, the actual value which the fraction $\frac{v_x}{v}$ assumes for any position of *p* may at once be obtained. Here *PL* represents the velocity of *p*, or *v*, while *PM* and *PN* denote the horizontal and vertical components v_x and v_y respectively.

In the right-angled triangle *PML*, the angle *MLP* equals the angle θ, their sides being perpendicular, and accordingly

$$\frac{-PM}{PL} = \sin\theta,$$

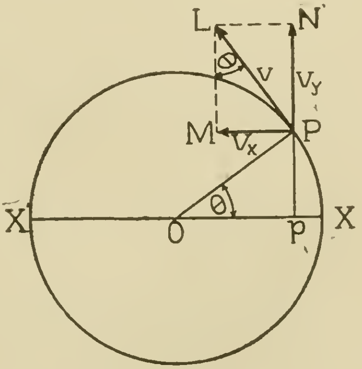

FIG. 44. Value of velocity in harmonic motion.

that is,

$$\frac{v_x}{v} = -\sin\theta,$$

or

$$v_x = -v\sin\theta.$$

From this it is seen that v_x, the velocity of p, is always a fractional part of v, the velocity of P, and, moreover, that it is such a fractional part as is represented by — sin θ. The minus sign here indicates that *the direction of the velocity and the value of sin θ are always of opposite sign.* That is, in the first and second quadrants, where sin θ is positive, the motion is from right to left, or negative; and in the third and fourth quadrants, where sin θ is negative, the motion is from left to right, or positive. Therefore, the velocity of p is a function of — sin θ, and varies as — sin θ varies. By giving proper values to θ, the velocity of p at any point of its path may be determined.

When $\theta = 0$ radians, $v_x = 0$,

$\theta = \frac{\pi}{2}$ radians, $v_x = -v$,

$\theta = \pi$ radians, $v_x = 0$,

$\theta = \frac{3\pi}{2}$ radians, $v_x = v$,

$\theta = 2\pi$ radians, $v_x = 0$.

This shows that at the extremity X of the path, the velocity of p is zero, that is, at this point p is at rest. As p moves to the left, its velocity increases in the negative direction and reaches a maximum at the center of the path, where it is equal in magnitude to the constant speed v of the particle P. As p moves farther to the left, its velocity decreases and becomes zero at X', where p is again at rest. As p now moves to the right, its velocity increases in the positive direction, and at the center O once more attains the maximum value v, equal to the constant speed of P in the circle. As p approaches X, its velocity again diminishes, and for a second time becomes zero at this point. As p continues thus to vibrate with harmonic motion, these variations in its velocity are repeated each period.

Inasmuch as

$$v = \frac{2\,\pi\,r}{T}, \qquad \text{(p. 51)}$$

and as

$$\theta = \frac{2\,\pi\,t}{T}, \qquad \text{(p. 62)}$$

the equation representing the velocity of a particle moving with harmonic motion may be written

$$v_x = -\frac{2\,\pi\,r}{T}\sin\frac{2\,\pi\,t}{T},$$

which is the usual form.

To here
from p. 60

64. The Acceleration in Harmonic Motion. Since the velocity of a particle moving with harmonic motion is continually changing, the motion of this particle is not uniform, but accelerated. Moreover, as the velocity is not only not constant in amount but even changes in direction, the acceleration as well as the velocity is variable. To determine this variable acceleration, we must again remember that, just as the *velocity* of the particle p, moving with harmonic motion, is the horizontal component of the *velocity of P* moving in the auxiliary circle, so also is the *acceleration of p* in its horizontal path the horizontal component of the *acceleration of P* in the circle.

The acceleration of P is that of a particle moving with constant speed v in the circumference of a circle of radius r; and this acceleration a has been shown (p. 55), to be constant in amount, namely,

$$a = \frac{v^2}{r},$$

and always directed towards the center of the circle. Accordingly, since acceleration is a vector quantity, we may *represent* the acceleration of P by the line a, Fig. 45, where the length of this line is proportional to the magnitude of the acceleration or $\frac{v^2}{r}$, and the direction of this line indicates the direction of the acceleration, which is towards the center of the circle. The acceleration a of the particle P may now be resolved into two components, one a_x in the horizontal direction, and the other a_y in the vertical direction. The horizontal component a_x is the acceleration at any instant of the particle p, as this moves with harmonic motion along the horizontal diameter, while the vertical component a_y is the acceleration at the same instant of a particle, likewise moving with harmonic motion, but along the vertical diameter.

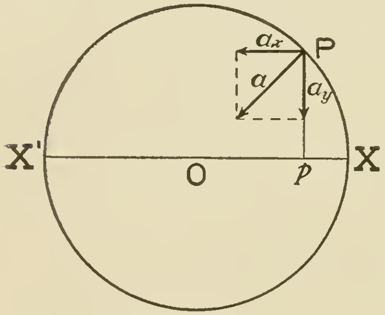

FIG. 45. Determination of the acceleration in harmonic motion.

By comparing with one another, the horizontal components of the acceleration of P at different points of the circle, we may deduce the law according to which the acceleration of p varies in its back-and-forth motion along the path. At X, Fig. 46, the acceleration of P being horizontal, the horizontal component of the acceleration is equal in magnitude to the acceleration itself. In other words, at this extremity of the path the acceleration of p is equal to the full acceleration of P, and is directed towards the position of equilibrium. As P approaches Y, the magnitude of the horizontal component diminishes. Therefore, although p is approaching O with an increasing velocity, its acceleration is decreasing. At Y the acceleration of P is directed vertically downward; and, as its horizontal component is then zero, the acceleration of p, as it passes through its position of equilibrium, is also zero. As P now moves from Y towards X', the horizontal component of the acceleration again increases, this time, however, in the opposite direction. At X' the horizontal component again attains the full value of the acceleration itself, since the acceleration of P is here once more horizontal. Similarly it is seen that, as P continues its circular motion from X' through Y' to X, the horizontal component of its acceleration diminishes as Y' is approached, becomes zero at Y', then increases, but once more with reversed sign, as P moves from Y' to X, and, finally, at X again becomes equal to its former maximum value. These variations, both in magnitude and direction, which the acceleration of p undergoes during a complete period, are represented in the accompanying vector diagram of accelerations, Fig. 47, and may be interpreted as follows:

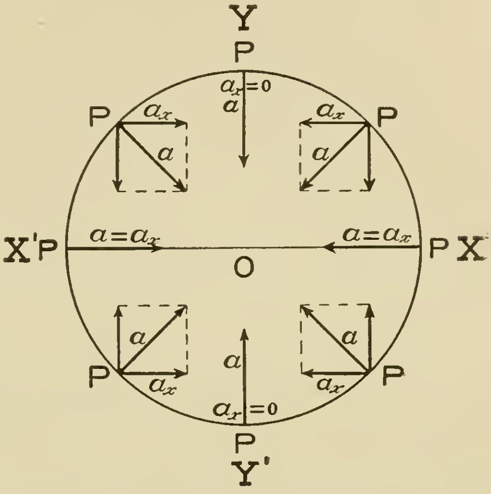

FIG. 46. The acceleration at different points of the path.

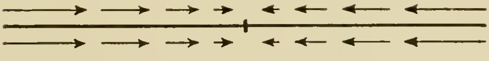

FIG. 47. Variation of the acceleration for an entire period.

At the extremities X and X' the particle has a zero velocity, that is, for an instant it is at rest. At the next instant it has begun to move, and so has a finite velocity. This change from a zero velocity to a finite velocity in this extremely short interval of time gives the maximum *rate* of change of the velocity and, therefore, the maximum acceleration. When the particle reaches the position of equilibrium, it has the velocity v, equal to the constant speed v of the particle P; and, accordingly, while it is passing this central point of its path, it is moving for an instant with uniform motion. For an exceedingly short time, therefore, there is no change in the velocity and hence there is no acceleration. Thus it is that the acceleration is greatest at the extremities of the path, where the velocity is zero, and is zero at the center of the path, where the velocity is greatest.

65. In Harmonic Motion the Acceleration is Proportional to the Displacement and Opposite to the Displacement in Sign. Thus it is seen that the acceleration of p is always directed towards the position of equilibrium, that it reaches its maximum value at the extremity of the path, where it becomes equal to the full acceleration of P, and that it becomes zero at the center. At any intermediate position its value may be shown, with the aid of Fig. 48, to be *directly proportional to the distance of p from the center*, that is *the acceleration is proportional to the displacement.*

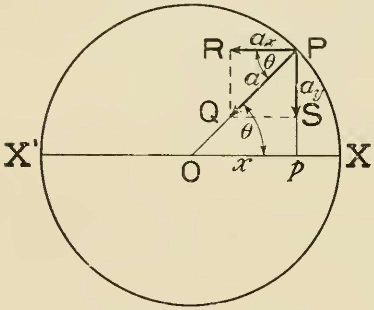

FIG. 48. The value of the acceleration in harmonic motion.

Here PQ, or a, represents the total acceleration of P, while PR and PS are the horizontal and the vertical components, a_x and a_y, respectively. The resolution of a into its components forms the triangle RPQ, which is similar to the triangle pOP. The sides of these triangles being proportional, we have

$$\frac{-a_x}{a} = \frac{Op}{OP}.$$

But Op is the displacement of p, here denoted by x, while OP is the radius r of the auxiliary circle. Accordingly, the above equation becomes

$$\frac{-a_x}{a} = \frac{x}{r},$$

or

$$a_x = -\frac{a}{r}x.$$

Since a is the acceleration of P, which is constant in magnitude, while r is the radius of the auxiliary circle, which also is constant, the value of the fraction $\frac{a}{r}$ is constant, a_x, being therefore equal to a constant multiplied by x, is proportional to x; that is, *the acceleration of p is proportional to the displacement.*

Furthermore, it is observed that this acceleration is directed towards the left when the displacement is to the right, or positive, and towards the right when the displacement is to the left, or negative. From this it follows that *the acceleration is not only proportional to the displacement in magnitude, but is also opposite to the displacement in sign.* Indeed, it is because the acceleration and the displacement are of opposite sign, that the minus sign must always occur in the above equation which represents the acceleration in harmonic motion.

The relations in harmonic motion here set forth, that the acceleration is always proportional to the displacement, and is always directed towards the position of equilibrium—that is, that the acceleration is positive when the displacement is negative, and negative when the displacement is positive —are characteristics so fundamental that their statement furnishes the most rigid definition of this motion. Accordingly, *simple harmonic motion* may be defined as motion in a straight line, in which *the acceleration at every point is proportional to the displacement in magnitude* and *opposite to the displacement in sign.*

66. Mathematical Expression for the Acceleration in Harmonic Motion. The relation existing between the acceleration PR, or a_x, of the particle p and the acceleration PQ, or a, of the particle P, Fig. 48, is given by the equation

$$\frac{-PR}{PQ} = \cos\theta,$$

or

$$\frac{a_x}{a} = -\cos\theta,$$

or

$$a_x = -a\cos\theta.$$

Here θ, or *RPQ*, is the phase angle, since it is equal to the alternate interior angle *XOP*. This expression shows that the acceleration of *p* is always a fractional part of the acceleration of *P*, and, indeed, it is the fractional part which is denoted by $-\cos\theta$. The minus sign is here used to indicate that *the direction of the acceleration and the value of cos θ are of opposite signs.* Since the acceleration of *p* is always directed towards the position of equilibrium, that is, towards the central point of the path, it follows that, in the first and fourth quadrants, where $\cos\theta$ is positive, the acceleration is negative, as it acts toward the left, while in the second and third quadrants, where $\cos\theta$ is negative, the acceleration is positive, as it acts here towards the right. Because the acceleration of *p* varies when θ varies, its value at any point depends upon the position of the particle *P*, which revolves in the auxiliary circle. Therefore, by giving different values to θ, we may determine the acceleration of *p* at different points of its path.

When $\theta = 0$ radians, $a_x = -a$,

$\theta = \frac{\pi}{2}$ radians, $a_x = 0$,

$\theta = \pi$ radians, $a_x = a$,

$\theta = \frac{3\pi}{2}$ radians, $a_x = 0$,

$\theta = 2\pi$ radians, $a_x = -a$.

From these values of a_x, it is evident that the acceleration of p reaches a maximum at the extremities X and X' of its path, and at these points, indeed, it is equal in magnitude to the full acceleration of P in the circle. The acceleration diminishes as p approaches the center, and becomes zero when the center is reached. As this point is passed, the acceleration changes its sign, increases again in magnitude and rises once more to the maximum a at the extremities X and X'.

In the expression, (p. 72),

$$a_x = - a \cos \theta,$$

if $\cos \theta$ is replaced by its value $\frac{x}{r}$, we have

$$a_x = - \frac{a}{r} x,$$

which is the equation for the acceleration in harmonic motion, as before derived (p. 71).

67. The period in Harmonic Motion. Besides furnishing the most rigid definition of harmonic motion, (p. 71), the equation.

$$a_x = - \frac{a}{r} x$$

is helpful in expressing the value of the period with which this motion takes place. Here a is the constant central acceleration of the particle P in the auxiliary circle, that is (p. 54)

$$a = \frac{v^2}{r},$$

so that

$$a_x = - \frac{v^2}{r^2} x.$$

In this equation, v is the linear velocity, or the speed, of P in its path, (p. 51), where

$$v = \frac{2\pi r}{T}.$$

The equation then becomes

$$a_x = -\frac{4\pi^2}{T^2}x,$$

where T is the *period*. This is the usual expression for the acceleration in harmonic motion.

When this equation is solved for T, or the period,

$$T = 2\pi\sqrt{-\frac{x}{a_x}},$$

or

$$T = 2\pi\sqrt{-\frac{\text{displacement}}{\text{acceleration}}}.$$

Inasmuch as the displacement and the acceleration are always of opposite signs, the quantity under the radical sign is always positive, and T is always positive and real. Moreover, since 2π is a constant quantity—a number—, and since the fraction

$$-\frac{\text{displacement}}{\text{acceleration}}$$

is always a constant, (p. 71), the value of T, depending as it does only upon constant quantities, is itself a constant. Accordingly, *the period in harmonic motion is perfectly constant*, and the vibrations of a particle executing harmonic motion take place in equal times. Furthermore, since the amplitude does not enter into this expression for the period, the all-important law is likewise deduced, that the period of a particle executing harmonic motion *is independent of the amplitude.*

The laws of harmonic motion here derived, namely, that *the acceleration is proportional to the displacement and opposite to the displacement in sign,* that *the period is independent of the amplitude,* and that *the vibrations take place in equal times,* are most fundamental, and are of frequent service in the study of the pendulum, and of problems in elasticity, sound, and light.

Tored

CHAPTER VII.

FORCE AND MOTION.

68. Definition of Matter. In the foregoing pages, we have considered motion simply in the abstract, without making inquiry as to its origin or as to the conditions under which it takes place. Moreover, we have regarded it from a purely geometrical standpoint, and have spoken of a moving particle or body, without giving further attention to the size or shape of this body, or to the amount, or quantity, of matter which is affected by the motion. In considering now the motion of a body as this actually occurs in the world about us, it is necessary, first to notice especially the *quantity of matter* the body contains, and later to take into account other attributes or conditions of this body.

The definition of matter previously given (p. 2), namely, that it is anything that occupies space, is far from complete because it does not tell what matter *is.* Since its nature at present is not well understood, matter is defined as above simply by one of its most obvious and distinctive qualities. If future investigation should reveal the nature of its structure and composition, it will in no way alter the fact that matter must occupy space; and so we may take this definition, which is obviously not incorrect, though inadequate, as only provisional and unsatisfactory.

69. Mass. Among the many ways in which bodies are found to differ in their characteristics, is the one which has to deal with their *masses*, that is, with the amounts, or quantities, of matter that enter into the constitution of these bodies. Accordingly, *mass* is defined as *the quantity of matter a body contains,* irrespective of the *kind* of material of which it is composed.

The mass of a body must be regarded as an inherent property of the body, and one that remains unchanged so long as matter is neither added to the body nor taken from it. Indeed, modern science is founded, in part at least, upon the doctrine of the conservation of matter, which assumes the constancy of mass and its independence of any change within the power of man. Thus the mass of a block of iron remains the same whether it is at one place or another on the earth's surface, or whether, in imagination, it be removed altogether from the earth, and carried to another planet or to interstellar space. The iron may even be melted or beaten into another form without affecting its mass, provided only that the iron remains the same iron, no particles of matter being added to it or taken from it.

70. Distinction between Mass and Weight. In forming the conception of mass noted above, we must free the mind completely from the more usual idea of the weight with which all matter is inevitably associated on the earth. As will be explained in detail later, the weight of a body is the attraction of the earth for the body; and so the weight of a body depends for its very existence upon the presence of the earth or some other body outside itself to attract it. The amount of matter the body contains, or its mass, is in no way affected by the presence or absence of this outside attracting body; hence its mass remains without increase or decrease, wherever in the universe the body may be situated. "The destructive effects of a cannon ball are due entirely to its mass and to the relative speed with which it impinges on the target, and would be exactly the same (for the same relative speed) in regions so far from the earth, or other attracting bodies, that the ball had practically *no weight at all*......... When we *open* a large iron gate properly supported on hinges, it is the mass with which we have to deal; if it were lying on the ground and we tried to *lift* it, we should have to deal simultaneously with its weight and with its mass." (Tait: Properties of Matter.)

71. Density. As regards their masses, bodies differ among themselves not only as to their volumes, but also as to the intensity with which matter fills those volumes. Therefore, the mass of a body depends both upon its volume and also upon the quantity of matter contained in unit volume; and so the entire mass of a body is the number of units in the volume of the body multiplied by the mass contained in unit volume. *The quantity of matter contained in unit volume is known as the density,* and this is always the ratio between the entire mass of the body and the number of units in the volume of the body. Accordingly,

$$\text{Density} = \frac{\text{mass}}{\text{volume}},$$

or

$$\text{Density} = \text{mass in unit volume.}$$

From this it is evident that *a body has unit density when unit mass is contained in unit volume,* that is, when the mass of 1 gram is contained in 1 cubic centimeter. The only body completely satisfying this definition of unit density is water, since 1 cubic centimeter of water has a mass of 1 gram. 1 cubic centimeter of iron, however, has a mass of 7.7 grams; 1 cubic centimeter of silver has a mass of 10.4 grams; 1 cubic centimeter of gold has a mass of 19.3 grams, etc. Therefore the density of water is 1; the density of iron 7.7; the density of silver 10.4; the density of gold 19.3, and so on.

72. Quantity of Motion. Since a body cannot move unless matter changes its position, the motion of a body implies that the matter contained in that body is imbued with a certain velocity. In all considerations of motion, therefore, it is necessary to regard not only the *mass* which moves but also the *velocity* with which it moves.

If two equal masses move with equal velocities, it is evident that the amount, or quantity, of motion in one, however this may be measured, is exactly equal to the amount or quantity of motion in the other. If the masses are now united, it is also evident that the quantity of motion in this two-fold mass is twice as great as that in either one of the single masses. Accordingly, if the velocity remains the same, the amount, or quantity, of motion in a moving body is directly proportional to the mass of that body.

Again, it is evident that, if a certain mass, moving with a certain velocity, contains a certain quantity of motion, the same mass, moving with a velocity twice as great, contains a two-fold quantity of motion. Therefore, if the mass remains constant, the quantity of motion in a moving body is directly proportional to the velocity with which the body moves. Since the quantity of motion in a moving body is thus proportional to the mass which moves and also to the velocity with which it moves, and to these alone, it must be proportional to their product, and be measured by their product.

73. Momentum. When motion is considered as a measurable quantity, the amount, or quantity, of motion is known as the *momentum*. The momentum of a moving body, therefore, is proportional to the product of the mass of the body, and the velocity with which the body moves; or

$$\text{Momentum} : m\, v,$$

that is,

$$\text{Momentum} = k\, m\, v,$$

where m is the mass of the moving body, v its velocity, and k the factor of proportionality.

74. The Unit of Momentum. No mention has yet been made of the unit, in terms of which momentum, or quantity of motion, may be measured, and the choice is a wide one because of the presence of the factor k. If we agree to call the momentum unity when the mass and the velocity are each equal to unity, then k must also be unity, that is

$$k = 1.$$

and in this case

$$\text{Momentum} = m\ v.$$

Therefore, in order that k may be equal to unity, *we define the unit momentum to be the momentum, or the quantity of motion, possessed by unit mass moving with unit velocity*, that is, by the mass of 1 gram moving with a velocity of 1 centimeter per second.

Moreover, since momentum is the product of mass and velocity, the magnitude of the separate factors may be varied at will, so long as the product remains constant. Thus, there is unit momentum, if a mass of 1 gram moves with a velocity of 1 centimeter per second, or if a mass of $\frac{1}{10}$ gram moves with a velocity of 10 centimeters per second, or if a mass of 10 grams moves with a velocity of $\frac{1}{10}$ centimeter per second, the product of mass and velocity being unity in each case.

75. Force. Observation of the conditions existing in the universe leads to the inevitable conclusion that the natural state of matter, that is, its condition of perfect equilibrium, is either that of complete rest or of uniform motion in a straight line. There is nothing within matter itself that gives to it the power of altering its state of rest or of uniform rectilinear motion. Experience shows that if a mass originally at rest is made to move, or if matter originally in motion is brought to rest or has its motion in any way changed, some action outside the body itself must have taken place, and this action is what is known as force.

In accordance with this conception, *force may be defined as any cause that alters or tends to alter a body's natural state of rest or of uniform motion in a straight line.* This simply means that force is an effort of any kind which produces or tends to produce motion in matter, or which changes or tends to change the existing motion of matter. A pull, a push, a blow, an attraction, a repulsion are examples of force. When matter at rest, however, is made to move, or when matter in motion is brought to rest or has its motion in any way changed, a change has taken place in the velocity, and an *acceleration* has been imparted. Accordingly, *force is any cause that imparts acceleration to matter.*

76. Measurement of Force. Since matter can be accelerated only by the action of a force, it is necessary to investigate the relation between the force which acts, the mass of matter upon which it acts, and the acceleration which it thus imparts.

If two forces act upon two equal masses of matter and produce an equal acceleration in each, it is evident that the forces themselves must be in every way equal. If the two equal masses are now united, and a force acts upon this double mass to produce the same acceleration, it is likewise evident that this must be a two-fold force. Accordingly, we see that when *different forces* act upon *different masses* to produce *the same acceleration the forces must be proportional to the masses.* On the other hand, if *two unequal forces* act upon *two equal masses* and produce an acceleration in one *twice as great* as in the other, the force producing the double acceleration is of twice the intensity of the one producing the single acceleration. Therefore, when *different forces* act upon *equal masses* to produce *different accelerations, the forces must be proportional to the accelerations.* Since forces are thus proportional to the *masses* upon which they act and to the *accelerations* which they impart, they must be proportional to the product of these two quantities, or

$$\text{Force} : \text{mass} \times \text{acceleration},$$

that is

$$\text{Force} = k \text{ mass} \times \text{acceleration},$$

or, more briefly,

$$\mathrm{F} = k\, m\, a,$$

where F is the force, m the mass, a the acceleration, and k the factor of proportionality.

77. The Unit of Force—The Dyne. Inasmuch as the above relation always holds true, we may derive from it a unit in terms of which it is possible to measure and compare forces. In order that the unit of force so derived may be an absolute unit (p. 18), it must be expressed in terms of absolute units only; and in this event k must equal unity. Accordingly, if we choose our unit so that F shall be unity, when m and a are each equal to unity, then in the above expression,

$$k = 1,$$

and we have

$$F = m\,a.$$

In giving to k the value of unity in this expression, we have, however thereby determined the absolute unit of force to be the force which will impart to 1 absolute unit of mass 1 absolute unit of acceleration. If we select, as the absolute units, those of the C. G. S. system, the unit of force thus determined is called the *dyne*. Accordingly, *the absolute C. G. S. unit of force, or the dyne, is the force which will impart to 1 gram of matter an acceleration of 1 centimeter per second in a second.*

A conception of the magnitude of the dyne may be obtained as follows: When 1 gram of matter falls freely under the *attraction of gravity* (p. 34), it acquires an acceleration of 981 centimeters per second in a second, while, as we have learned above, *the dyne* acting on 1 gram of matter imparts to it an acceleration of only 1 centimeter per second in a second. Therefore, *the force of gravity acting on 1 gram of matter—that is, the weight of 1 gram—is a force of 981 dynes;* and, accordingly, one dyne is a force equal to $\frac{1}{981}$ part of the weight of 1 gram of matter.

78. The Acceleration Imparted is Inversely Proportional to the Mass.

In the expression

$$F = m\,a,$$

we see that force is equal to the *product* of mass and acceleration. For a given force, therefore, the factors may vary as much as we please, so long as the product remains constant. If the mass upon which the force acts is very large, the acceleration imparted must be correspondingly small; and conversely, if the mass is small, the acceleration must be large in the same proportion. Thus, a force of 1000 dynes acting upon a mass of 1 gram will impart 1000 units of acceleration; but a force of 1000 dynes acting upon a mass of 1000 grams will impart only 1 unit of acceleration. Moreover, a force of 1000 dynes acting upon a mass of 2 grams will impart 500 units of acceleration; while a force of 1000 dynes acting upon a mass of 500 grams will impart 2 units of acceleration, and so on. From this we see that the acceleration imparted by a given force is always inversely proportional to the mass.

79. Masses Compared by a Comparison of the Accelerations Produced by a Force. Since the accelerations imparted to bodies by a given force are inversely proportional to the masses of the bodies, it is possible, by comparing these accelerations, thereby also to compare the masses. This may be accomplished, approximately, at least, by connecting by a coiled spring two small carriages, Fig. 49, which roll freely upon a smooth table. If two unequal masses are placed upon these carriages and the latter drawn apart, the tension in the spring will cause equal forces to act upon each, and the masses will be accelerated. If, for instance, the masses are 2 kilograms and 1 kilogram and if, for the purpose of illustration, the masses of the carriages may be neglected, the acceleration imparted to the 2 kilo-

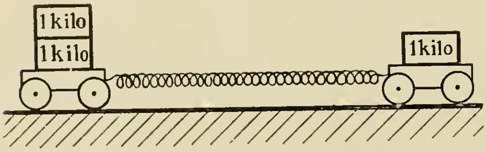

Fig. 49. A fundamental method of comparing two masses.

gram mass will be only half as much as that imparted to the 1 kilogram mass. Since the accelerations imparted are proportional to the distances passed over in a given time (p. 33), we shall find that the distance passed over in a certain time by the 2 kilogram mass is only half as great as that passed over in the same time by the 1 kilogram mass.

If, now, the experiment is repeated, using, however, an unknown mass instead of the 2 kilograms, we shall find that the accelerations are again unequal, from which we may conclude not only that the masses are unequal, but that these masses are to each other inversely as their respective accelerations. The ratio of the accelerations may be found, as before, by measuring the distances passed over by each mass in a given time, and the masses are then inversely as these distances. Thus, if the unknown mass passes over 10 centimeters, while the 1 kilogram passes over 50 centimeters, then

$$\text{Unknown mass} : 1 \text{ kilogram} = \frac{1}{10} : \frac{1}{50},$$

or,

$$\text{Unknown mass} = 5 \text{ kilograms}.$$

This method of comparing masses is thus seen to be absolutely independent of weight, and would operate equally well in regions so far from the earth that the masses had practically no weight at all.

80. Force Expressed in Terms of Mass and Rate of Change of Velocity. It is possible, however, to look at force from a slightly different point of view. Remembering that acceleration is the rate of change of velocity, or that

$$a = \frac{V - V_0}{t}, \qquad \text{(p. 29)}$$

we may substitute this value of *a* instead of *a* itself in the fundamental expression for force,

$$F = m\,a,$$

and obtain

$$F = m\,\frac{V - V_0}{t}.$$

Here $V - V_0$ is the change that has taken place in the velocity of the mass m in the time t, so that $\frac{V - V_0}{t}$ denotes the *rate* of change of this velocity, due to the action of the force F. We may therefore express force as the product of the mass and the rate of change of velocity.

From this expression for force, we may likewise derive a definition of the unit force. Inasmuch as there is no factor of proportionality, we see that if $V - V_0$ is unity, and if m and t are unity, F also becomes unity. Accordingly, the unit force is the force that in unit time will impart to unit mass unit velocity; or, adopting the C. G. S. units, *the unit force, or the dyne, is the force that in 1 second of time will impart*
~~The expression of~~ force may also be written — to the mass of 1 gram
a velocity of 1 cm. per. sec.

81. Force Expressed as the Rate of Change of Momentum.

The above expression for force may also be written

$$F = \frac{m\,(V - V_0)}{t},$$

or

$$F = \frac{m\,V - m\,V_0}{t}.$$

Here the numerator of the fraction represents the change of the momentum that has taken place in time t, and the fraction itself denotes the *rate* of change of momentum. Therefore, the above expression defines force as *the rate of change of momentum.*

From this expression, also, we may derive a definition of the unit force. If the numerator is made equal to unity, that is, if $m\ V - m\ V_0$ represents the change of *unit* momentum and if t likewise equals unity, then F will equal unity. Accordingly, *the unit force is the force that in unit time will impart unit momentum.*

These three definitions of the unit force are equivalent, and may be used indifferently.

82. Impulse of a Force. By rewriting in slightly different form the last expression for force, we obtain

$$F\ t = m\ V - m\ V_0.$$

That is, the product of the force and the time during which the force acts equals the change of the momentum produced. The product of the force and the time during which it acts is known as the *impulse* of the force, that is,

$$F\ t = \textit{impulse}.$$

The effectiveness of a force in producing or changing motion may be expressed by saying that *the impulse of a force always equals the change of momentum imparted.*

Since impulse is a *product*, of which the factors are force and time, it follows that the time during which a force acts is just as effective as the force itself in producing a change in the momentum of a body. That is, a small force acting for a very long time may produce a considerable change in the momentum; whereas even a great force may act for a time so exceedingly short that the impulse is small, and the change in the momentum is correspondingly small.

Since mass enters into every expression for force, and since the effect of a force is to change the velocity of this mass, the word ''motion'' which occurs in the definition of force, (p. 79), must be extended so as to mean momentum. Therefore, a more nearly correct definition of force would be: *force is any cause that produces or tends to produce a change in the momentum of a body.*

If the body is at rest before the force acts upon it, then

$$V_0 = 0,$$

and the above expressions for force reduce to the following:

$$F = m\,a,$$

$$F = m\,\frac{V}{t},$$

$$F = \frac{m\,V}{t}.$$

where V denotes the final velocity of the body, and therefore *the total change in the velocity due to the action of the force.*

83. The Characteristics of a Force. The characteristics of a force are those attributes which specify or differentiate one particular force from any other. These characteristics are three in number; the *direction,* the *intensity,* and the *point of application.*

The *direction* of a force is the characteristic which *gives direction to the momentum* imparted by the force. The change of momentum produced by the action of a force always takes place in the direction in which the force acts.

The *intensity* of a force is the characteristic which determines *the amount of the momentum* developed in a certain time by the action of the force. The amount of the momentum so developed is always directly proportional to the intensity of the force.

The *point of application* of a force is characteristic which has to do with the *quality, or kind, of motion* communicated to an extended body by the action of a force upon it. This motion may be one of translation, or one of rotation, or both, depending upon the point of the extended body at which the force is applied. The point of application may always be displaced by any amount in the direction in which the force acts, without changing the effectiveness of the force. A push, therefore, is in every way equivalent to a pull, both being in the same direction and of equal intensity.

84. Composition and Resolution of Forces. Inasmuch as force is a directed quantity, it is evident that when two or more forces act at the same time upon the same point, they may be replaced by a single force, the effect of which will be the same as the combined effects of the several forces. This single force is called the resultant, and the several forces that constitute this resultant are called the components. Accordingly, *the resultant of a system of forces is the single force that can produce the same effect as the several forces acting, and therefore it is the single force that can take the place of the several forces.*

If *two equal and opposite* forces act upon a particle, each force holds the other in equilibrium, the resultant is then zero, and the particle remains at rest. Accordingly, a particle is said to be *in equilibrium,* that is, it has no motion imparted to it by the existing forces, when the resultant of all the forces acting upon it is zero. *A force equal to the resultant and acting in the opposite direction,* since it holds the resultant in equilibrium will also hold the system of forces in equilibrium and may be called the *anti-resultant.*

When forces are parallel, whether they are in the same or in opposite directions, the resultant may be found by taking their algebraic sum. If forces meeting at a point make an angle with one another, they may be compounded according to the "parallelogram law," where the resultant of two forces is the diagonal of the parallelogram founded upon these two forces as sides. This resultant may be compounded in the same way with another force, and so on until all have been used. Moreover, any force may be regarded as a resultant, and may be resolved into its components in any two directions by the general law for the resolution of vectors previously given (p. 41).

85. The Triangle of Forces. If three forces are in equilibrium, that is, if the resultant of three forces is zero, the three forces must *meet at a point,* they must *lie in one plane,* and they must be represented in intensity and direction by *the three sides of a triangle taken in order;* that is, each force must be equal and opposite to the resultant of the other two. This is known as the triangle of forces.

The truth contained in the principle of the triangle of forces may be illustrated as follows: *AB* and *AC*, Fig. 50, are two forces acting at point *A*, *AD* is their resultant, being the diagonal of the parallelogram founded upon *AB* and *AC* as sides. In order to hold these two forces in equilibrium, that is, in order to have a zero resultant, it is necessary to add the third force *AD′*, which is the anti-resultant of the other two. Since *AD* and *AD′* are in equilibrium, it follows that the three forces, *AB*, *AC*, and *AD′* are also in equilibrium. Accordingly, Fig. 51, if *AB* represents one force in intensity and direction, and *BD*, which is the equal of *AC*, represents the second force, then the side *DA* will also represent in intensity and direction the force *AD′*, which holds the forces *AB* and *AC* in equilibrium. The forces *AB*, *AC* and *AD′*, which are in equilibrium, may therefore, be represented in direction and intensity by the three sides of the triangle, *AB*, *BD*, and *DA*, *taken in order.*

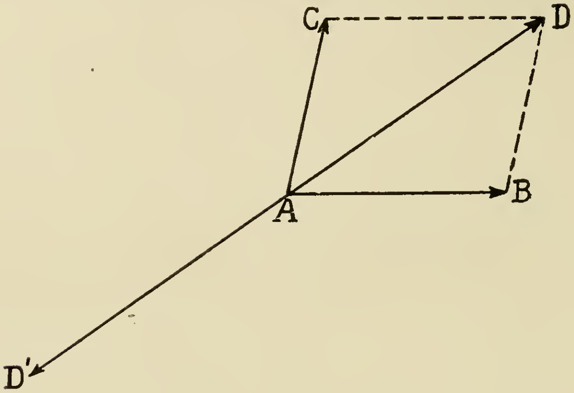

FIG. 50. Equilibrium of three forces.

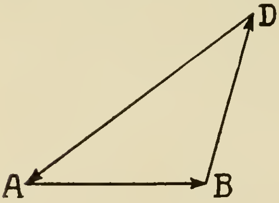

FIG. 51. The triangle of forces.

Any number of forces *in equilibrium* may be represented in intensity and direction by the sides of a closed polygon. This polygon need not be plane.

A ball supported on a jet of water beautifully illustrates the action of three forces in equilibrium. The impact of the water upon the ball is a force that may be represented by *F*, Fig. 52, the weight of the ball acting vertically downward is a force represented by *W*, while the water flowing along the side of the ball produces a diminution in the pressure of the air on that side. Accordingly, there is an unbalanced pressure of the air, that is, a force, represented by *P*, acting upon the opposite side. If these three forces may be represented by the three sides of a triangle *taken in order*, that is, if each of these forces is equal and opposite to the resultant of the other two, the resultant is zero, and the ball remains in equilibrium supported upon the jet.

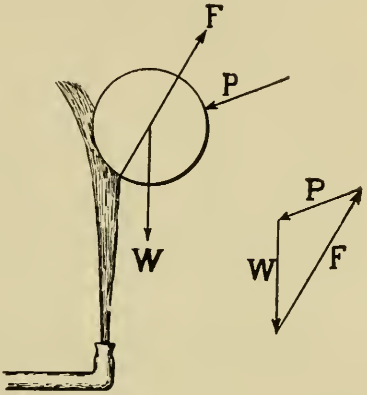

FIG. 52. Ball on a jet of water illustrating three forces in equilibrium.

86. Moments. We have heretofore considered a force chiefly as regards its effect in producing translation, that is, causing motion in a straight line, where the direction of the motion coincides with the direction of the force, and the quantity of motion is proportional to the intensity of the force. We have seen, however, that when a force acts upon an extended body, the force may produce a motion of translation or one of rotation, depending upon the point at which the force is applied. If the extended body is perfectly free, and the force acts in the direction of a point within the body known as the *center of mass*, the motion is one of pure translation, and the particles of the body, move forward in lines which are straight and parallel. If, however, the force acts upon the body in a direction other than towards the center of mass, the effect of the force is to produce not only a translation but also a rotation.

When an extended body is not free, but is so supported at any one of its points that it can rotate about that point, a force, unless it be directed toward the point of support, will cause the body to rotate, and the force is said to give rise to a certain *moment*. *The moment of a force is the measure of the importance of that force in producing rotation,* and depends not only upon the *direction* and *intensity* of the force, but also upon *its point of application.*

The point about which the rotation takes place is called the *center of moments.* The perpendicular distance between the center of moments and the force, or between the center of moments and the direction of the force produced, is called the lever arm of the force. The longer the lever arm, the greater is the "leverage" of the force, and the greater is its effectiveness in producing rotation. *The moment of a force, therefore, is measured by the product of the intensity of the force and the length of the lever arm.*

Thus, if the shaded portion, Fig. 53, represents any body whatever, free to rotate about the point O, the point O is called the center of moments. The force F, acting at the point a, will produce rotation about O, since O does not lie in the direction AF in which the force acts. The effect of the force upon the body will remain the same if the point of application of the force is changed from a to b or c, since both of these points lie in the direction of the force. The perpendicular distance Op between the center of moments and the direction of the force is the lever arm of the force. This distance, then, is one of the two factors that determine the moment of the force, the other being the intensity F. Accordingly, the effectiveness of this force in producing rotation, that is, *the moment* of the force F about the point O is

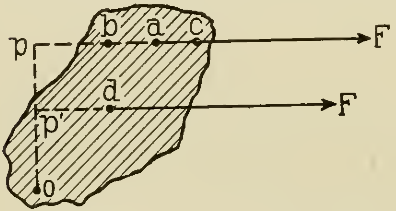

Fig. 53. The principle of moments.

$$F \times Op.$$

If the point of application of the force had been d instead of a, b, or c, the lever arm of the force would have been Op', and the moment would have been only

$$F \times Op'.$$

When a force produces rotation, the rotation is regarded as positive or negative according to a purely arbitrary convention. Rotation in the direction of the hands of a watch, or clockwise, is considered negative, while rotation in the opposite direction, or counter-clockwise, is positive.

87. Conditions of Equilibrium for a Rigid Body. According to the location of the point of application, a force may produce a motion of simple translation, or one of rotation, or perhaps both. When several forces act at different points upon an extended body, the resulting motion is the combined effects of all forces acting. If the body is to be *in equilibrium*, that is, if there is to be no motion whatever, there must be neither a resultant force to produce translation, nor a resultant moment to produce rotation. In other words, there are *two conditions that must be fulfilled if equilibrium is to result: the algebraic sum of all the forces in any direction must be zero, and the algebraic sum of the moments about any axis must also be zero.*

88. A Rigid Body in Equilibrium. The conditions for the equilibrium of a rigid body under the action of several forces may be illustrated in Fig. 54. Here the forces F and F' equal to 1 and 2 units respectively, act at points 10 and 40 of the graduated bar, which for the sake of simplicity we may suppose to be without weight. Since these forces are parallel and in the same direction, the resultant R is equal to their sum, or 3 units. A resultant, however, cannot hold a system of forces in equilibrium; and the effect upon the bar is exactly the same as if

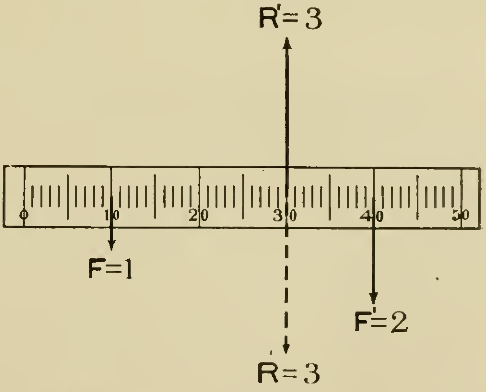

FIG. 54. A body in equilibrium under the action of several forces.

the forces F and F' were removed and the resultant R operated in their place. Accordingly, if *equilibrium* is to be produced, we must apply, as an anti-resultant, a force R', equal to the resultant and in the opposite direction. In doing this, we have complied with the first of the two conditions for equilibrium; for the algebraic sum of the forces F, F' and R' is zero.

As yet nothing has been said as to the location of the resultant R or of the anti-resultant R'. This, however, may be determined by remembering that, as the resultant may in every way take the place of the forces acting, it must also produce the same amount of rotation, that is, it must have a moment equal to the sum of the moments of F and F'. If we choose the point O as the center of moments, although any other point might be taken equally well, we see that the moment of F about O is

$$1 \times 10, \text{ or } 10, \text{ (negative rotation)},$$

where 10 is the lever arm; and the moment of F' about O is

$$2 \times 40, \text{ or } 80, \text{ (likewise negative rotation)},$$

where in this case 40 is the lever arm.

The algebraic sum of the moments of F and F' about O is, therefore,

$$10 + 80, \text{ or } 90, \text{ (negative rotation)},$$

so that the moment of the resultant R must also be 90, producing negative rotation. But the intensity of R is 3; and its lever arm, which we may call x, must be of such length that

$$3 \times x = 90,$$

from which

$$x = 30.$$

Accordingly, if the resultant *R* were to replace the forces *F* and *F′*, it would have to be applied at point 30, in order that its moment might be equal to the sum of the moments of *F* and *F′*. Therefore, if we are to establish the condition of *equilibrium*, the anti-resultant *R′* must be applied at point 30 in order that its moment about *O* may be 90 and produce rotation in the positive direction.

From this we see that the forces *F* and *F′* applied at points 10 and 40 respectively, and the force *R′* applied at point 30 will produce equilibrium; for the algebraic sum of these forces is zero, and the algebraic sum of their moments is likewise zero.

89. The Couple. Two forces, equal in intensity and opposite in direction, acting at different points upon an extended body, constitute a couple. A couple, therefore, consists of two equal, parallel forces, acting in opposite directions, but not at the same point nor in the same line.

Since the two forces of a couple are always equal, parallel, and opposite, their resultant is zero as regard translation, and the entire effect of a couple is to produce rotation. Wherever the point may be about which the rotation takes place, that is, whatever point may be chosen as the center of moments, the moment of the couple is the algebraic sum of the moments of the two forces about that point. Thus, Fig. 55, if *O* is the center of moments and *Oa* and *Ob* are the lever arms of the two forces respectively, the moments of the left hand force about *O* is

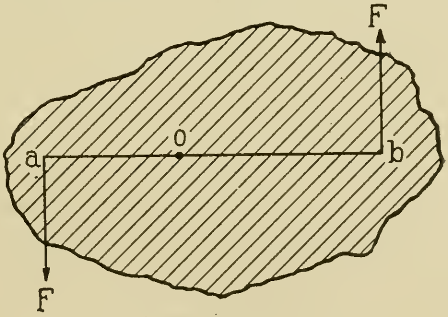

FIG. 55. The couple.

$$F \times Oa, \text{ (positive rotation)},$$

and the moment of the right hand force about this point is

$$F \times Ob, \text{ (also positive rotation)}.$$

Since both forces tend to produce rotation in the same direction, the moment of the couple is the sum of the moments of the two forces, or

$$\text{Moment of the couple} = F \times Oa + F \times Ob,$$

or

$$F\ (Oa + Ob),$$

or

$$F \times ab.$$

If the center of moments is taken on the line *ab* produced, or, indeed, anywhere else, either within or without the body, the moment of the couple remains unchanged. *The moment of a couple is thus seen to be the product of one of the forces and the perpendicular distance between them,* and to be wholly independent of the position of the center of moments. Since the resultant of a couple is zero as regards translation, no single force can replace a couple. A couple may be held *in equilibrium* only by the action of another couple of equal moment tending to produce rotation in the opposite direction.

90. Equilibrium of Couples. The freedom from rotation, which may be attained only by the action of one couple holding another in equilibrium, may be illustrated by placing two couples of equal and opposite moments upon the bar shown in Fig. 56. The two forces F, each equal to 2 units, and the two forces F', each equal to 1 unit, constitute the two couples, the perpendicular distance between the forces being 10 for the first couple, and 20 for the second. Since the moment of a couple is the product of one of the

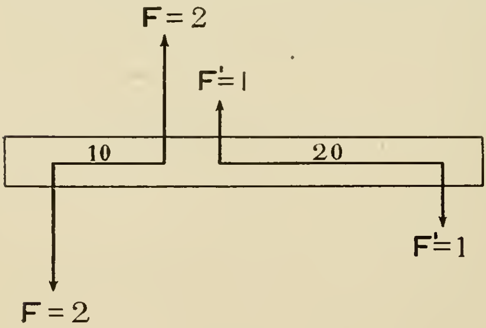

FIG. 56. Two couples in equilibrium.

forces and the perpendicular distance between them, we see that the moment of the first couple is

$$2 \times 10, \text{ or } 20, \text{ (positive rotation),}$$

and that the moment of the second couple is

$$1 \times 20, \text{ or } 20, \text{ (negative rotation).}$$

The algebraic sum of these moments is therefore zero; and as the algebraic sum of the forces is likewise zero, the body or bar is completely in equilibrium (p. 91).

From this we see that a couple cannot be held in equilibrium by a single force, but only by the action of another couple of equal moment tending to produce rotation in the opposite direction.

Omit from here

91. Center of Mass. We have seen, (p. 90), that the moment of a force with respect to a certain center of moments is the product of the intensity of the force and the perpendicular distance let fall from the center of moments upon the force. By an extension of this principle, we may define *the moment of mass with respect to a plane* to be the product of the mass and the perpendicular distance let fall from that mass upon the plane.

Thus, Fig. 57, the moment of the mass m with respect to the plane $ABCD$ is the product of the mass m and the perpendicular distance r between this mass and the plane, that is,

$$\text{Moment of mass } m \text{ with respect to plane } ABCD = mr.$$

In a similar way.

$$\text{Moment of mass } m' \text{ with respect to plane } ABCD = m'r'.$$

If perpendicular distances on one side of the plane are considered as positive, these distances on the other side must be taken as negative, so that the moment of mass m'', situated on the other side of the same place and at a distance r'' from it, is negative, or

Moment of mass m'' with respect to plane $ABCD = m''r''$.

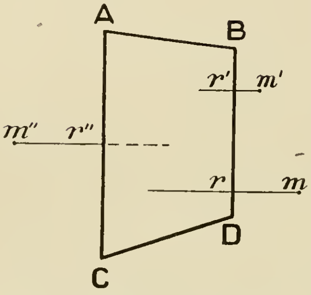

FIG. 57. The moment of mass with respect to a plane.

If m, m', and m'' are now considered simply as particles of an extended body, through which the plane $ABCD$ passes, it is evident that the moment of mass of each of the other particles of the body may be determined in a similar manner. It is further obvious that, in general, according to the location of the plane, some of these moments of mass will be positive and others negative, and that the resultant moment of mass of the entire body with respect to this plane will be the algebraic sum of all the moments of mass of the individual particles. *If the plane is now moved parallel to itself until the resultant moment of mass of all the particles is zero,* that is, if the sum of the moments of mass on one side of the plane is exactly equal and opposite to the sum of the moments of mass on the other side, the plane is then said to pass through the point which is known as the *center of mass* of the body.

If this is true for any plane as $ABCD$, it must be true for any other plane, and therefore it is true for all planes. Accordingly, if different planes are passed in all possible directions through the body, and each plane is moved parallel to itself until the resultant moment of mass with respect to this plane is zero, each plane must then pass through the center of mass of the body, and this point is therefore determined as the common point in which all these planes intersect. From this it follows that *the center of mass* of an extended body may be defined as that point within the body such that, if a plane be passed in any direction through this point, the resultant moment of mass of all the particles of the body with respect to this plane is zero. This does not mean necessarily that there are just as many particles of given mass on one side of the plane as on the other, since these masses may be at widely different distances from the plane on the two sides, but it *does* mean that the *moments* of these masses are equal and opposite.

Thus, for instance, if we assume that the masses of the various particles are all equal, it means that two particles at a certain distance from the plane on one side will have a moment which is equal and opposite to that of one particle at twice this distance on the other. Therefore, in order that the moments of unequal masses on the two sides of the plane may be equal and opposite, the masses must be to each other inversely as their distances from the plane, as may be seen in Fig. 58. Here, if M and M' represent two different masses connected by a slender rod, and if O is the center of mass, the moment of the mass M about O is equal and opposite to the moment of the mass M' about the same point, that is,

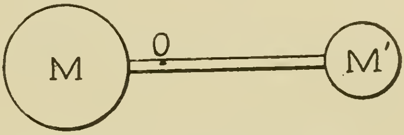

FIG. 58. The center of mass.

$$M \times MO = M' \times M'O,$$

and

$$\frac{M}{M'} = \frac{M'O}{MO}.$$

The center of mass is thus seen to divide the distance between the two masses inversely proportional to the masses.

From the location of the center of mass, we see that, if the various particles of a body were acted upon by parallel forces, each force being proportional to the mass of the particle upon which it acted, the algebraic sum of the moments of these forces with respect to the center of mass would be zero. In this case, the resultant of this system of forces, which would then be their sum, would have its point of application at the center of mass of the body.

The moon does not revolve about the earth as if the latter were fixed in space, but both revolve about their common center of mass, this being a point nearly 3,000 miles from the center of the earth. If a stick is thrown into the air, although it may rotate or gyrate apparently in an irregular manner, its center of mass, if the resistance of the air is neglected, will nevertheless describe a parabola as perfect as if the stick consisted of but a singe particle and this were located at the center of mass. If a shell is fired from a cannon and bursts in the air, however widely and irregularly the pieces may fly, the center of mass of these fragments will continue in the same parabolic path in the air after the explosion as the shell itself would have pursued had the explosion not taken place.

In dealing with the action of a force upon an extended body, it is often convenient to consider a body as a single point, and to regard the entire mass of the body as concentrated at its center of mass.

92. Conservation of Momentum. In all of the above instances, a force from within the system of bodies itself acts upon the various parts of the system, thereby imparting to them a momentum. This force, in the case of the earth and the moon, for instance, is of the nature of a tension which acts upon one as strongly as upon the other; and in the explosion of the shell it is of the nature of a pressure which acts in one direction upon one fragment just as strongly as it acts in the opposite direction upon another. The momentum given to a fragment in one direction, therefore, is exactly equal to that communicated to another fragment in the opposite direction, so that the resultant momentum imparted to all the fragments of the shell is zero, and the center of mass remains undisturbed. We see, therefore, that it is only a force exterior to the body or the system of bodies that can produce any change in the resultant momentum.

These examples serve to illustrate the general principle that, when the parts of a system react upon one another, the momentum gained by one part is exactly counterbalanced by the momentum lost by another. In other words, the total momentum of a system can be neither increased nor decreased by any interaction between its parts. This doctrine is one of general application in mechanics, and is known as the principle of the conservation of momentum.

93. Inertia. Important among the various properties of matter in claiming our attention is the one which relates to the behavior of matter of all kinds when compelled by a force to change its condition of rest or of uniform motion in a straight line. Because we observe that objects in motion ultimately lose their motion and come to rest, we are accustomed to think of rest as the natural state of matter. A closer examination, however, into the conditions existing in nature leads to the inevitable belief that uniform motion in a straight line is a condition as natural to matter as is rest itself. Moreover, if matter could be wholly removed from the action of outside forces, it would continue forever in the state of rest or of uniform motion in a straight line in which it happened to be at the instant when it was freed from all external influences. Matter, therefore, has no power within itself of changing its state of rest or of uniform motion in a straight line; and this property of maintaining unchanged its existing condition as regards motion, is known as inertia.

It is because of inertia that a person standing in a car is thrown backward when the car starts, and forward when it stops. It is also because of inertia that a person running swiftly finds it impossible to stop at a certain point, but is carried past that point before he can bring himself to rest. It is due to the inertia of matter that danger arises from a collision on the railway. The train with its contents cannot be brought immediately to rest when the engine strikes the obstacle, but continues to move forward, the tendency thus to continue moving being greater than the strength of the materials of which the cars are made; and so, in consequence of their inertia, the cars are crushed and flattened against the obstacle and against one another before they can be brought to rest. It is due to inertia that the dust may be removed from a carpet by beating. Under the influence of the blow the carpet moves away from the dust, while the latter by its inertia remains at rest and is blown away by the wind.

94. Resistance to the Action of a Force due to Inertia. These and many other illustrations serve to show that when matter is compelled by a force to change its state of motion, its inertia causes it to offer a certain opposition to that change, and this opposition acts as a resistance to the force. It is to be noticed that this resistance is offered only to a *change* of motion and not to motion itself, and therefore, it is a resistance that is offered, not to a velocity but to *an acceleration.*

On this account a person standing upon a spring balance in an elevator will apparently weigh more as the elevator starts to ascend, and less as it starts to descend. In the first instance, the acceleration of gravity acting on the person is increased by the acceleration of the elevator, and in the second instance it is decreased by the same amount. When, the acceleration has ceased, however, that is, while the elevator is moving upward or downward with constant velocity, the person's true weight will be given by the indication of the balance.

95. Masses Measured by their Inertia. This resistance, due to inertia, which is necessarily present when a force acts upon a body to change its motion, is always equal in intensity and opposite in direction to the force acting. If the body to be moved is freed from all obstacles to motion, such as forces acting in the opposite direction, friction, resistance of the air, etc., so that a force acting on a body has only the inertia of the body to overcome, the force will set the body in motion, however small the force or however large the body. Moreover, if this force acts with constant intensity, the motion of the body will be uniformly accelerated, and the acceleration imparted will be inversely proportional to the mass of the body, as must follow from the equation, (p. 81),

$$F = ma.$$

In other words, a large mass has a great inertia, and, therefore, for a given force the acceleration imparted is comparatively small; whereas a small mass has a relatively small inertia, and hence, under the action of the same force it acquires a correspondingly great acceleration. Masses, therefore, are proportional to their inertia and may be measured by their inertia.

It is because the inertia of the 2-kilogram mass, Fig. 49, is twice as great as that of the 1-kilogram mass, that the resistance of the former to the action of the spring is twice as great as that of the latter. The difficulty of changing its condition of motion being twice as great for two kilograms as for one, the velocity which the two-fold mass acquires in a given time is only half as much as that which the single mass attains in the same time. In other words, the acceleration which the tension of the spring gives to the 2-kilogram mass is only half as much as it imparts to the 1 kilogram mass. The masses, therefore, are proportional to their resistances to change of motion, or to their inertia, that is, they are inversely proportional to the accelerations which a given force will impart.

96. The Influence of Time upon the Effect of a Force. When the motion of a body changes, the change of momentum, as we have seen, (p. 85), is equal to the impulse of the force that produces it, that is,

$$F\,t = m\,V - m\,V_0.$$

From this it is obvious that the change of momentum, depending as it does upon the product Ft, depends quite as much upon the time t as upon the force F. In other words, a change of momentum cannot take place instantly, for it requires the perfectly finite time t for the force F to produce the given change. Because of this, *inertia may be defined as that property of matter in virtue of which a finite time is required for matter to change its velocity.*

Illustrations of this principle abound. An untrained horse attached to a heavy wagon will dash forward, possibly breaking his harness, and will produce at most only a jolt in the wagon, the inertia of the latter preventing its acquiring in so short a time the velocity of the horse. If a railway train attempts to start too suddenly, the wheels of the locomotive turn uselessly around and produce but little motion in the train, because the mass of the train, on account of its inertia, requires a much longer time to acquire a high velocity under the action of the given force.

Even though the intensity of the force may be relatively great, if this force acts upon a body for an exceedingly short time, the t in the above expression may have so small a value that the product Ft, or the *impulse* of the force, is necessarily very small. From this it follows that the velocity imparted to the body is correspondingly small, and, indeed, in many instances is so minute that the body is not observed to move at all.

97. Illustrations of Inertia. Let us consider, for example, a heavy ball, Fig. 59, suspended by a string just strong enough to hold it, while a piece of the same string is attached to the ball below. A steady pull upon the lower string will break the string above the ball, while a sudden jerk will cause it to break below the ball, the resistance of the particles of the ball to so great an acceleration in the latter case being greater than the strength of the string.

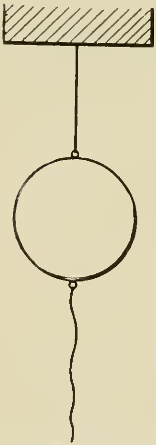

FIG. 59. The inertia of a lead ball.

In this way a candle may be shot through a board, or a cannon ball may be fired through a door standing ajar without perceptibly moving the door. Here the impact of the candle against the board or of the cannon ball against the door is of so short duration that the force of cohesion of the wood is much smaller than is the force that would be required in the same short time to give the board or the door the velocity of the ball, and so the wood breaks and gives way before the board or the door as a whole acquires an appreciable velocity. In a similar way, a lead bullet when fired from a gun is crushed upon itself as it begins to move, and thus fits into the rifling of the gun. The flight of a bird depends upon the inertia of the air. The stroke of the wing, if it were made slowly, would cause the air immediately underneath to move. The stroke, however, is of a duration so short that in this small interval of time the air acquires very little velocity indeed, and hence the bird is supported upon the air, which is quite rigid under the almost instantaneous blow.

A most remarkable illustration of the inertia of the air is exhibited in tests of gun-cotton made in the United States Navy.

This material is produced in the form of compressed cylindrical cakes, Fig. 60, in which the letters U. S. N., and the date are stamped to a depth of two or three millimeters. A hole is likewise left in the center in which the detonator is to be placed. When this cake of gun cotton is laid face downward upon a flat iron block and exploded, the upper surface of the block is hollowed out, Fig. 61, a deep hole is forced in the center, and the letters and the date in reverse are found to be deeply stamped in the iron. Here the duration of the explosion is so short, and the particles of air in the letters and in the central hole are given so high a velocity that a far greater force would be required in this exceedingly short time to check the velocity of these air particles than is required to tear apart the molecules of the iron. Accordingly, the iron yields, and the particles of air are literally projected into the metal, the air, under the velocity with which it is moving, being actually harder than the iron.

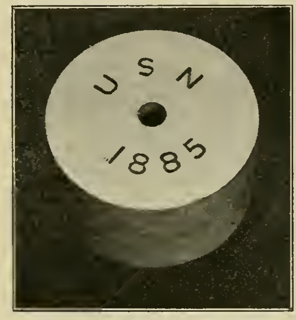

FIG. 60. Gun-cotton used in the U. S. Navy.

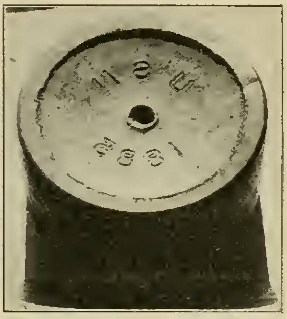

FIG. 61. During the explosion of gun-cotton, the air in the letters is driven by its inertia into the block of iron upon which the explosive is placed.

It is a curious fact that so frail an object as the leaf of a tree placed between the gun-cotton and the iron will partly shield the latter from the violence of the explosion, and a raised impression of the leaf is left upon the surface of the iron. In this case, the leaf, because of its inertia, cannot acquire in so short a time a velocity as great as that of the gases due to the explosion. While, therefore, the entire surface of the iron is rendered concave, the parts immediately beneath the leaf are less deeply impressed than are the adjacent portions; and so the iron retains an impression of the leaf, not stamped into the metal, but raised in relief from its surface, every vein and every detail showing as clearly as in the original leaf.

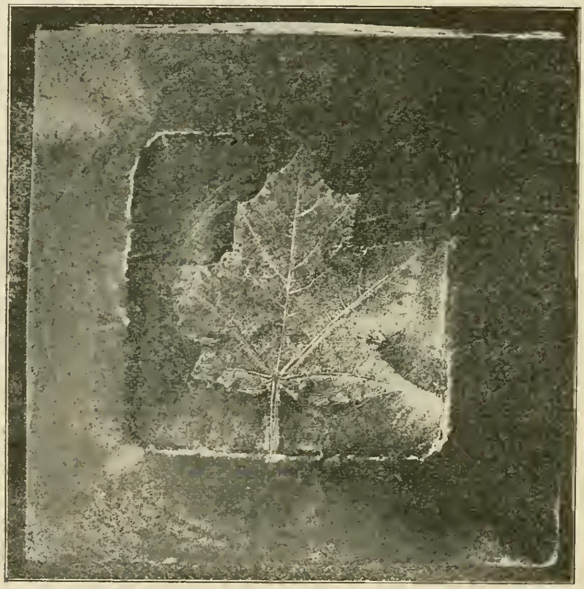

FIG. 62. A leaf, by its inertia, partially shields a block of iron from the explosion of gun-cotton.

98. Force Developed by Inertia—The So-called Force of Inertia. The converse of this principle is likewise true. Just as a force must act upon a body for a finite time to impart motion to the body, so also must a force act for a finite time and through a definite distance in order to stop a body when in motion. Even a small force, accordingly, if allowed to act for a sufficiently long time, will effectively bring a moving body to rest. On the other hand, if the motion of the body is *suddenly* arrested, the inertia of the body will offer a resistance to this sudden change of velocity, and this will operate as a force, or a blow, delivered by the moving body upon the opposing obstacle. It follows that the more quickly a body is stopped, the greater will be the force so developed. Upon this principle depends the action of the hammer, the pile driver, etc.

99. Newton's Laws of Motion. All the knowledge the world possesses regarding motion and the action of force in producing or changing motion is contained in the statements enunciated by Newton more than two centuries ago, known as the laws of motion. These laws were then announced with such clearness and with so remarkable a comprehension of the conditions existing in motion, that we use them today practically as they were given. As to a proof of these laws, it may be remarked that they are so because it is observed that all motions take place in strict accordance with them. We have, then, no direct proof that they are true, except as we perceive that every motion obeys these laws, and, therefore, every existing motion may be said to be a proof. These laws are thus a summing up of the results of experience and observation, and represent the conditions that obtain throughout the universe as regards motion and the action of a force in producing or changing motion.

While it is true that Newton announced these laws and that they are called by his name, the real credit for their discovery belongs not to Newton but to Galileo, who was the first to understand the true nature of motion. In all of his teaching as well as his writing, Galileo applied these laws to the movement of bodies, and thus he may be said to be, indeed, the founder of the science of mechanics.

These laws are three in number, and are as follows:

1. Every body continues in its state of rest or of uniform motion in a straight line, unless it is compelled by a force to change that state.

2. Change of momentum is equal to the impulse of the force that produces it, and takes place in the direction of the force.

3. To every action there is an equal and contrary reaction. or the mutual actions of two bodies are always equal and opposite.

100. The First Law of Motion. The first of these laws is known as the law of inertia, and expresses simply what is meant by inertia. It is evident that a body at rest will remain at rest until a force outside itself sets it in motion. It is, however, not so apparent that a body in motion will continue to move with uniform motion in a straight line until it is stopped by a force, since it is impossible upon the earth to free a body entirely from the action of forces that ultimately bring it to rest. The more completely, however, these opposing forces are eliminated, the more nearly does the motion of the body approach the state of uniform motion when no force acts, as described by Newton. Thus a stone thrown along a road is soon brought to rest by friction and by collisions with various objects. The same stone will travel much farther if projected along the surface of smooth ice, and will then only slowly change its motion as it is gradually brought to rest by the very slight friction between the ice and itself. We can see, therefore, that if the ice were perfectly smooth, that is, offering no friction whatever, and if no other forces opposed its motion, the body would then move forward forever with uniform motion in a straight line.

The existence of motion, therefore, does not necessarily imply the present action of a force, because uniform motion in a straight line is possible only when no force whatever acts. If a body moves with uniform motion in a straight line without the action of any force, it is evident that at some previous time a force has acted to impart this motion, and has then ceased to act. The body, therefore, because of its inertia, continues to move with uniform motion in a straight line with whatever velocity and in whatever direction it was moving at the instant when the action of the force ceased.

101. The Second Law of Motion. As regards the second law of motion, it would seem that this law is derived from the equation for force given above (p. 101), since it is a statement in words of just what this equation expresses. It must be remembered, however, that this equation, together with all that has been said, and, indeed, all that we know about the action of a force in producing motion, is a summation derived from experience. Before the time of Galileo, it was thought that one force could operate only after the action of a previous force had altogether ceased, and so it was imagined, for instance, that a bullet fired from a gun could fall to the ground only after the "onward" force of the bullet, as it was called, was completely exhausted. Galileo saw, however, what is now universally known to be true, that the effect of a force upon a body is independent of the previous history of that body as regards its motion. A generation, then, before the time of Newton, he expressed this fact; and, indeed, a part of the truth contained in the second law of motion is found in the statement, known today as Galileo's Principle, namely: Every force produces its own effect whether it acts alone or in composition with other forces. It was in accordance with this principle that Galileo conceived the possibility of combining several forces into one resultant; and it was he who first applied the principle of the composition of forces to the true explanation of the motion of a body under the combined action of several forces.

102. The Third Law of Motion. The third law of motion expresses the fact that under no conditions can a force exist singly, but that a force is invariably only one of a pair of forces that always exist together, and are always exactly equal and opposite. Instead, therefore, of a single force acting to produce motion, we must henceforth recognize the fact that there are always two forces to be considered, one exerted by the first body upon the second, while the other is exerted by the second body upon the first. It is this second force, or this other one of the pair of forces, that is meant by the word "reaction" in the law given above.

103. Discussion of the Third Law of Motion. A body lying on the table exerts a downward force upon the table equal to the weight of the body. It is just as true, however, that the table exerts an equal upward force upon the body, so that the body is in complete equilibrium under the action of two absolutely equal and opposite forces.

A horse drawing a canal boat is pulled backward by the boat precisely as much as the boat is pulled forward by the horse. In other words, the "action" of the horse is exactly equalled by the "reaction" of the boat, and this action, together with the reaction, constitutes the tension in the rope between them. The horse and the boat are thus under the action of strictly equal and opposite forces, and are completely in equilibrium as regards each other, because neither moves with respect to the other. It may be thought that the horse must pull forward *just a little* harder than the boat pulls backward, inasmuch the horse sets the boat in motion. This, however, is an error, since if it were true, the horse would then have a motion relative to the boat, that is, he would increase the distance between the boat and himself. The explanation lies in the fact that the horse, the rope, and the boat must be regarded as one system, and that a force is exerted by the horse's feet not against the rope or the boat, but against the earth, moving the latter backward; and it is a component of the reaction of the earth against the horse's feet that pushes the system forward.

A sky-rocket ascends because of the action of a force which is simply the reaction of another force. According to its size, the rocket contains a quantity of powder varying in mass from a few grams to one or two kilograms, all of which, in the gaseous condition, is driven forth with great velocity. The burning powder exerts a pressure which expells the gases thus formed, that is, it imparts an enormous acceleration to a quantity of matter equal to that of the unburned powder, and in so doing it exerts a force which may be many times greater than the weight of the rocket. This force, however, acting downward and accelerating the gases, is but one of the pair of equal and opposite forces which are always present, the other being a force, which is likewise greater than the weight of the rocket, acting upward. It follows, therefore, that, because the rocket is acted upon by a force greater than its own weight and in the upward direction, it rises, and continues to rise until this upward force ceases to act; that is, until the powder is burned, and the force required to accelerate the burned gases, to which the upward force is equal and opposite, no longer exists.

When the powder explodes in a gun, a force is developed between the bullet and the gun; and the force that acts upon the bullet moving it forward is only one of the two forces present, the other being the force acting upon the gun, moving it backward. This latter force, moreover, is just as great as the former, and imparts just as great a momentum to the gun as the former force imparts to the bullet. The recoil of the gun, therefore, is just as necessary as is the motion of the bullet; and since the momentum of the gun backward is exactly equal to the momentum of the bullet forward, the velocity of the gun is as much less than the velocity of the bullet as the mass of the gun is greater than the mass of the bullet.

The equality of the momentum of the gun backward and of the bullet forward is another illustration of the principle of the conservation of momentum (p. 98); for here the resultant momentum is zero, as must be the case when the members of a system produce motion by any interaction between their parts.

104. The Centripetal Force. When a body moves uniformly in the circumference of a circle, we have seen, (p. 51), that it is moving at any point of its path in the direction of the tangent at that point. That is, for an exceedingly short time the motion of the body in the arc of the circle coincides with the direction of the tangent; and, as the body is moving with constant speed in the circle, it moves for an instant with uniform motion in a straight line, according to Newton's first law. At the next instant, however, the body has changed its direction of motion, and moves in the direction of the tangent at a succeeding point. From the fact that the body is not moving continuously with uniform motion in a straight line, as would be the case if no force acted, it is evident that a force is operating upon the body to change its direction of motion. Moreover, the change in the direction of motion is away from the tangent and towards the center of the circle; and hence, by Newton's second law, the force that produces this change must likewise act towards the center. Furthermore, since the curvature of the circle is uniform, this change is constant in amount in equal succeeding intervals of time, and hence the force that acts continuously to deviate the body from the direction of the tangent is likewise constant.

The intensity of this force is obtained at once from a consideration of the acceleration towards the center, which we have already determined (p. 54). We have seen that, when a body moves uniformly in the circumference of a circle, there is a constant acceleration, a, acting towards the center, where

$$a = \frac{v^2}{r}.$$

Since acceleration is one of the factors that enter into the equation of force, mass being the other, it follows that we shall derive the expression for the intensity F of this force if we multiply both sides of the equation by the mass m of the moving body. Accordingly,

$$m\,a = F = \frac{m\,v^2}{r}.$$

From this it is seen that, when a body of mass m moves uniformly in the circumference of a circle, there is a constant force F, acting towards the center and continually changing the direction of the velocity, where

$$F = \frac{m\,v^2}{r}.$$

It is to be observed that this force, which is called the centripetal force, is always directed towards the center of the circle, and therefore it always acts at right angles to the direction of motion. No component of this force acts in the direction in which the body is moving, and so the motion of the body along its path is neither increased nor diminished by the action of this force. Accordingly, the entire action of the centripetal force is expended against the inertia of the body in changing its direction of motion, and causing it continually to move towards the center away from the tangent in which it tends to move by Newton's first law.

105. The Centrifugal Force. The application of the third law of motion, however, shows that this centripetal force is but one of a pair of exactly equal and opposite forces; the other, called the centrifugal force, being found in the resistance, or reaction, of the body to the action of the centripetal force. The centripetal force is a force that emanates from the center and acts upon the body, drawing it towards the center. The centrifugal force, therefore, is a force that emanates from the body and acts upon the center, as if it would draw the center towards the body. This reaction to the centripetal force, or the centrifugal force, is due solely to the inertia of the body, that is, to the resistance which the body offers to the change of motion produced by the centripetal force. The centrifugal force, therefore, is a true force, although it has existence only as the reaction of the centripetal force, and is to be sought, not in the action of the center upon the body, but in the action of the body upon the center. This centrifugal force (fleeing from the center) is wrongly named, for it is not a force that attempts to carry the body away from the center, but is the resistance which the body offers to being moved nearer to the center away from the straight line in which it tends to move.

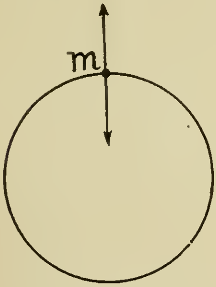

FIG. 63. The centrifugal force is equal and opposite to the centripetal.

The centrifugal force is always exactly equal and opposite to the centripetal force, Fig. 63, and the two together constitute the tension in the medium that connects the body with the center. This medium may be very apparent, such as the string by which a body is whirled about the hand; or it may be more intangible, though just as truly having existence, as, for instance, the force of gravitation, which constantly draws the earth towards the sun, thus causing the earth to move in its orbit instead of passing off in a straight line into space. The centrifugal force can never cause a body to move away from the center, because as soon as the centripetal force disappears (the string breaks, etc.), the reaction to this force, or the centrifugal force, likewise disappears, and the body then remains in motion in the direction of the tangent without the action of any force.

Familiar illustrations of the action of the centrifugal force abound. The water flies off from the periphery of the grindstone when the centripetal force, and therefore, also the centrifugal force, is greater than the force of cohesion of the water. The water, moreover, is observed to leave the grindstone, not in the direction of the radius, that is, away from the center, but in the direction of the tangent in which it moves as soon as freed from the action of the centripetal force. A body may be whirled about the hand by a string so long as the force (the centripetal force) required to draw the body from the direction of the tangent is not greater than the strength of the string. When the centripetal force becomes greater than this amount, the string breaks, the centrifugal and centripetal forces vanish together, and the body, by its inertia, moves in the direction of the tangent at whatever point it happens to be at that time. A vessel of water may likewise be whirled in a vertical circle without spilling when the centripetal force is greater than the weight of the water. In this case, the centrifugal force is also greater than the weight of the water, that is, the resistance offered by the water to the change of direction of its motion is greater than its own weight. The earth is not spherical, but is in the form of an oblate spheroid, due to the fact that as it revolves upon its axis the portions near the equator have a higher velocity, and consequently a greater centripetal force, than those near the poles. The reaction of this force is also greater, and accordingly, when the earth was plastic, the more effective resistance at the equator to motion in a circle caused the equatorial diameter to exceed the polar by about twenty-six miles. If a spherical vessel containing water and mercury is made to revolve rapidly upon the whirling table, Fig. 64, the liquids will be seen to form a band about the center, the mercury being nearer the glass, that is, farther from the center than the water. This follows at once from the fact that the mass of the mercury is greater than that of the water, and hence the reaction to the change of the direction of motion is greater for mercury than for water. A chain which has been set in rapid rotation by being mounted upon a vertical wheel, will, if slipped off, roll along the table like a hoop, each link endeavoring by Newton's first law to continue moving in a straight line, and offering a resistance to departure therefrom which makes the chain quite rigid.

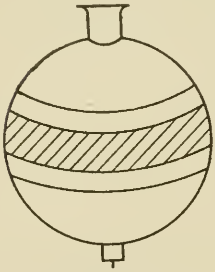

FIG. 64. The action of the centrifugal force upon water and mercury.

Practical applications of this tendency are found in the centrifugal separator in dairies and in the centrifugal drier in laundries. For many years, also, this principle has been employed in the construction of the centrifugal governor for the automatic regulation of steam engines. This device consists of two heavy balls suspended by rods so hinged at the top of a vertical rotating shaft that they are compelled to rotate with the shaft as the latter turns. When the system is at rest, the balls assume the lowest possible position; but when rotation takes place, they offer by their inertia great opposition to motion in a circle, and this tendency to motion in a straight line increases with the velocity of rotation. As the motion becomes greater, accordingly, the balls swing out from the axis, and move in larger and larger circles. A sliding collar on the rotating shaft is connected with the balls and also with the valve admitting steam to the engine, so that the steam may be shut off when the balls fly out at too great an angle, that is, when the engine is running too fast.

106. Elasticity. In the previous treatment of the action of a force upon a body, it has been assumed that the molecules of the body are relatively fixed in position, and therefore do not move among themselves. These molecules are believed to be particles that are not in contact, but separated from one another by spaces, or pores, the latter being relatively very large as compared with the size of the particles themselves. Because of the cohesion of matter, it is seen that molecular forces must act across these inter-molecular spaces, holding the molecules in their relatively fixed positions. It is found possible, however, by the application of an external force, to crowd these particles a little more closely together, or to pull them a little farther apart. When the positions of these particles are disturbed, it is also found that forces come into play tending to restore them to their original positions with reference to one another. If the molecules have their relative positions changed through the action of an external force the body is said to be deformed, or distorted.

That property of matter, in virtue of which a body tends to regain its original shape and size after it has been deformed by a force, is known as elasticity. The force exerted by the molecules after deformation, tending to restore the body to its former condition, is called the force of restitution; and in bodies that are perfectly elastic, this force is exactly equal to the force applied to the body to deform it. For every body there is a limit, beyond which it is impossible to carry the distortion without occasioning a permanent set; and after this limit has been exceeded, the particles refuse to return to their original positions. This limit, although varying for different substances, is perfectly definite for each material, and is known as the elastic limit. The ability of a substance to be distorted through very wide ranges without exceeding the elastic limit is known as elastic toughness. This property, however, is not to be confounded with elasticity, for the degree of elasticity indicates the completeness with which the particles return to their original positions after distortion has taken place. Thus, rubber has greater elastic toughness than steel, though steel is more nearly perfectly elastic than rubber.

There are several ways in which a body may undergo deformation, to each of which there belongs a separate kind of elasticity. Thus, there is elasticity of form, elasticity of volume, elasticity of bending, elasticity of extension and compression, elasticity of torsion, or twisting.

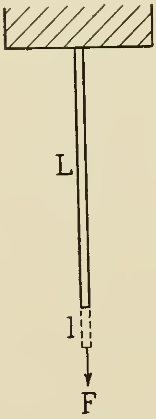

FIG. 65. Elasticity by stretching—Hooke's law.

It is found by experiment that when forces act upon a body to stretch it, *the elongations produced are proportional to the forces applied*, this relation being known as Hooke's law. Investigation shows that the amount l that a wire stretches, Fig. 65, is proportional to the length L of the wire, and to the applied force F, and is inversely proportional to the cross-section Q of the wire. The amount of elongation is also proportional to a constant k, where k depends upon the material of which the wire is made, and may be called the "ability to stretch." Therefore,

$$l = k\frac{FL}{Q}.$$

Quiz Y 12/18 p. 115

from which

$$\frac{1}{k} = \frac{F\,L}{l\,Q}.$$

The quantity $\frac{1}{k}$,which is always denoted by the letter E, is called the modulus of elasticity; and accordingly,

$$E = \frac{F\,L}{l\,Q}.$$

If the wire is of unit length ($L = 1$), and of unit cross-section ($Q = 1$), and if the elongation is unity ($l = 1$), then the modulus of elasticity E will be equal to the applied force F. The modulus of elasticity, therefore, or Young's modulus, as it is sometimes called, is the force that will stretch a wire of unit cross-section to twice its original length.

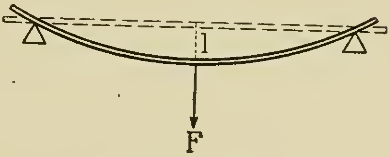

FIG. 66. Modulus of elasticity by bending.

Quiz

One other kind of elasticity may be considered, namely, that of bending, or flexure, and from this Young's modulus may likewise be obtained. If the body is in the form of a bar, Fig. 66, and rests upon two supports, one at each end, and if the force acts at the center tending to bend the body, then

$$\mathrm{E} = \frac{1}{4}\,\frac{F\,L^3}{l\,b\,h^3}.$$

where F is the force applied, L the length of the bar between the supports, l the amount of bending at the center, b the breadth, or width, of the bar, and h the height, or thickness, of the bar. The derivation of this expression is too difficult to be given here.

CHAPTER VIII.

WORK AND ENERGY.

107. Work. When a force acts upon a body either to produce motion or to change the magnitude of the existing motion, the force is said to do work. In other words, if, during the continuous action of a force, the body moves in the direction of the force, or in the direction of a component of the force, work is performed. Accordingly, work is that which is accomplished by a force that causes the body upon which it is directed to move; and the amount of work performed is proportional to the intensity F of the force and to the distance, or length l, through which the body moves while the force acts. If the force makes an angle α with the direction in which the motion takes place. Fig. 67, it is only the component of the force which lies in the direction of motion, or $F \cos \alpha$, that is effective in performing work, and in this case

$$\text{Work} : F\, l \cos \alpha.$$

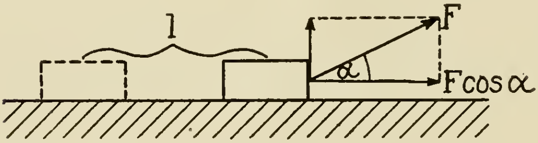

FIG. 67. A component of the force is effective in doing work.

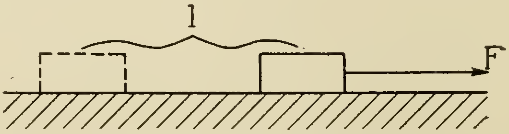

FIG. 68. The total force performing work.

If, however, the direction of the force coincides with the direction of motion, Fig. 68, the angle between the two is zero, and

$$\text{Work} : F\, l,$$

or

$$\text{Work} = k\,F\,l.$$

108. The unit of Work, The Erg. From the last expression it is possible to derive a unit in terms of which work may be measured; and if the quantities which specify this unit are expressed by absolute units only, the unit of work will likewise be an absolute unit. Accordingly, if we agree to call the work unity when the force and the distance are equal to unity, then

$$k = 1,$$

and the above expression becomes

$$\text{Work} = F\,l.$$

In thus giving to k the value of unity in the expression for work, we have thereby defined the absolute unit of work to be the work which is accomplished when one absolute unit of force acts through one absolute unit of distance; and if we choose as our absolute units those of the C. G. S. system, in which the absolute unit of force is the dyne and the absolute unit of length is the centimeter, we have the absolute unit of work known as *the erg*. *The erg,* therefore, or the absolute unit of work in the C. G. S. system, *is defined as the work which is performed by the force of 1 dyne acting through the distance of 1 centimeter.*

109. The Factors in Work. It must be remembered that work is a product, the factors of which are force and distance. It is necessary, therefore, that the point of application of the force should move through a certain distance in the direction of the force in order that work may be done. No work is performed by forces in equilibrium, because in this case no motion is produced. Neither is work done when a force acts at right angles to the direction of motion. Thus, no work is accomplished by the centripetal force in uniform circular motion, since no component of this force lies in the direction in which the motion takes place. For the same reason the force of gravitation, though holding the earth in its orbit and causing it to revolve around the sun, performs no work. The force of gravity performs no work upon a body at rest; but work is done upon the body as soon as the latter falls, for then a force equal to the weight of the body moves through a distance.

Furthermore, since a given amount of work depends upon the product of two factors, these factors may be varied at will, so long as their product remains constant and equal to the given number of units of work. Thus, 100 ergs of work may be performed by a force of 1 dyne acting through 100 centimeters, or by a force of 2 dynes acting through 50 centimeters, or by a force of 25 dynes acting through 4 centimeters, or by a force of 100 dynes acting through 1 centimeter, and so on. It should be noticed that *time does not enter into the conception of work,* and that *the amount of work done is independent of the time required to perform it.*

While both of the factors that enter into work, namely, force and distance, are directed quantities, it is of interest to observe that work, which is the product of these two quantities, is itself wholly independent of direction, and therefore is not a vector quantity.

110. Multiples of the Erg—The Kilogrammeter and the Joule. Since the dyne is an exceedingly small force, the erg, which is the work done by the force of one dyne acting through the distance of one centimeter, is a very small quantity of work indeed. On this account it is often convenient to make use of a unit of work larger than the erg, and the unit frequently adopted is the kilogrammeter. This is the work done by a force equal to the weight of one kilogram acting through the distance of one meter. Since gravity acting on 1 gram of matter exerts a force of 981 dynes, the force due to gravity acting upon 1 kilogram is 981,000 dynes. When this force acts through 1 meter, or 100 centimeters, 981,000 × 100 ergs of work are done. 1 kilogrammeter of work, therefore, represents an amount of work equal to 98,100,000 ergs.

Another multiple of the erg, namely, 10,000,000 ergs, or 10^7 ergs, is very generally used, and is called the joule.

111. Power of Activity. While the element of time is not contained in the expression for work, it does enter into the conception of the *rate* at which work is performed, and this is known as power, or activity. Accordingly, power, or activity, is defined as *the rate of doing work,* that is, the amount of work done in the unit of time, or the second. If the rate at which work is performed is constant, this is equivalent to saying that power, or activity, is the ratio of the entire amount of work performed to the entire time spent in accomplishing this work, or

$$\text{Power, or activity} = \frac{\text{work}}{\text{time}}.$$

If the unit of work, the erg, is performed in the unit of time, the second, the power, or activity, is likewise unity. The absolute unit of power, therefore, is the ability to perform work at the rate of 1 erg per second. Since the erg, as we have seen, is a very small unit, the absolute unit of power is so exceedingly small that a multiple by 10,000,000 is usually taken as a practical, or working, unit, and is called the watt. The watt, therefore, equals 10^7 ergs per second, or 1 joule per second. It requires 746 watts to equal the unit of power frequently used, known as the horse power.

112. Energy. When work is performed upon a body, as, for instance, when a stone is lifted vertically from the ground, the work which is done is equal to the product of the force expended and the distance through which the force has been overcome. As a result of this expenditure of work, the stone no longer has its place upon the ground, but is in an elevated position. From the very fact that the stone is raised, it is evident that as soon as it is released, the force of gravity will again cause it to move through the same distance, and an amount of work will thus be done *by the stone,* equal to the amount of work that was done upon it in giving it its elevation. Since the stone before falling is in a position to do work, it is said to possess energy, where *energy is defined as the ability to perform work.* The elevated position of the stone indicates that work has been done upon it, and this work is represented by its position of advantage with reference to the earth. As long as the stone maintains its position, the work that has been done upon it to give it this elevation may be conceived as stored up in it, and the energy it possesses may be thought of as the work that the stone can do, and sometime will do, but has not yet done. *Energy therefore, may be defined as stored-up work, or work that has not yet been performed.*

113. Potential Energy. What is true of the stone when lifted from the earth, is likewise true of any body whatever upon which work has been expended. The work done upon a body is never lost, but remains stored up in the body to be given out again when the body once more resumes its original condition. The energy possessed by a body in virtue of its position, as described above, is but one of the many forms in which this quantity appears, and is known as *potential* energy. Since energy is the result of work performed, and since it may again be converted into work, *energy and work are the same in nature, and both are measured in ergs.*

114. Kinetic Energy. When a body is set in motion by the action of a force, the point of application of the force must move through a certain distance and work must be done, that is, energy must be expended, in overcoming the inertia. A moving body, therefore, possesses energy, because work must be performed in setting it in motion. Moreover, since a moving body requires a force to stop it, and since a velocity cannot be changed instantly, the force required to stop the body must act through a certain distance; and so work must be done by an exterior force before the body can be brought to rest. A moving body, therefore, possesses the ability to do work while it itself is losing its motion; and its energy of motion, accordingly, must be equivalent to the work done by the force in stopping it. The energy possessed by a moving body is known as *kinetic* energy.

If a mass m, moving with a velocity v, is brought to rest by a force F, acting through the distance l, the work done by the force in stopping the body is

$$\text{Work} = F\,l.$$

But (p. 81)

$$F = m\,a,$$

and (equation 8, p. 33)

$$l = \frac{v^2}{2\,a};$$

therefore

$$\text{Work} = F\,l = m\,a\,\frac{v^2}{2a} = \frac{m\,v^2}{2}.$$

But the expression *Fl* represents the work done in stopping the body, and the equivalent expression $\frac{m\,v^2}{2}$ represents the ability of the moving body to perform this amount of work, hence its kinetic energy. Therefore, the energy of a moving body, or its

$$\text{Kinetic energy} = \frac{m\,v^2}{2}.$$

It should be noticed that the kinetic energy of a body depends only upon the mass of the body and upon the square of its velocity, and is independent of the acceleration of gravity. Hence the kinetic energy of a moving body would be the same for the same velocity wherever in the universe it might be situated.

115. The Conservation of Energy. The store of any particular kind of energy possessed by a body seldom remains constant but undergoes frequent transformation, and reappears in one or another of the many forms in which energy is capable of existing. The change that is continually taking place from potential energy to kinetic and back again to potential is beautifully illustrated by the swinging of the common pendulum. Work is done upon the pendulum in lifting it against the force of gravity, and the equivalent of the work thus expended is represented by the potential energy the pendulum possesses in its elevated position. At the lowest point of the swing, this energy of position has entirely disappeared, having been transformed into the energy of motion, which is here a maximum. Because of the kinetic energy which the pendulum then possesses in virtue of its motion, it is able again to lift itself against the force of gravity, and its energy of motion is once more transformed into energy of position. At a point intermediate between its highest and lowest positions, its energy is partly potential and partly kinetic; as the pendulum falls, its potential energy diminishes and its kinetic energy increases, but the sum of the two remains perfectly constant throughout.

If the principle here set forth is extended so as to apply to all systems in which energy of one kind is transformed into energy of another, it is evident that in the phenomenon of the swinging of the pendulum we observe the operation of a great law of the universe, in accordance with which energy of one kind is continually disappearing, only to reappear in another of the many forms in which energy exists. Of these various forms, *potential energy is the most highly available*, because it is also the most unstable; and so it is continually passing over into the kinetic form, or into one of the less useful manifestations into which energy may be changed. Indeed, it may be said, as a general principle of the universe, that *potential energy tends towards a minimum.*

Furthermore, if any number of particles form what is known as a closed system, that is, if these particles move among themselves due to their own mutual actions, but do not receive energy from without, it will be found that the energy of such a system can neither be increased nor decreased by the mutual interaction of its parts. One part loses in energy exactly as much as another part gains, and so the total energy of a closed system remains constant. If the closed system is extended so as to include the entire universe, the various objects which we recognize will then be parts of this closed system, and to the entire universe with all of its parts the same principle applies. This great law, which is known as the doctrine of the conservation of energy, reveals the important truth that *energy itself is indestructible;* and however varied may be the changes and transformations it may undergo, the sum total of the energy of the universe remains perfectly constant. It follows, therefore, that when energy of one kind disappears, it does so only to reappear unchanged in amount in some other form.

116. Machines. When a certain amount of work is to be performed, it not infrequently happens that the store of energy at hand is of a form unsuited to the accomplishment of this work, and that the energy can become available only by the intervention of some intermediate device. A mechanism for transferring energy from one point to another, and for transforming energy from one form to another to make it more adaptable for a particular purpose, is known as a machine. In all machines a force is exerted from the storehouse of energy upon some part of the mechanism, this part is then moved through a certain distance, and work is thus performed upon the machine. The machine transfers this work to another part, which in general exerts a different force upon the resistance, or obstacle, to be overcome, this resistance is also moved through a certain distance, and the machine performs work upon it.

Although a force exerted by the source of energy may be indefinitely multiplied by the action of a machine, the energy can never be increased, and therefore no more energy can be obtained from a machine than is originally supplied to it. There is, then, no such principle as the conservation of force, while in every machine, as elsewhere, the law of the conservation of energy is strictly obeyed. Not all of the energy supplied to a machine is available for the performance of useful work, since a portion of it is expended in overcoming friction, stiffness of cords, belts, etc. No machine, accordingly, is capable of performing a quantity of work equal to the energy expended upon it, and on this account perpetual motion is impossible. The ratio of the useful work performed *by* the machine to the energy supplied *to* it is called the *efficiency* of the machine.

However complicated a machine may be, it is found to consist in its essential features of one or more of the so-called "mechanical powers," which are effective in transferring and adapting energy to meet a particular need. These mechanical powers are six in number, and are as follows: the lever, the wheel and axle, the inclined plane, the screw, the wedge, and the pulley.

117. The Lever. The lever consists of a bar free to turn about a fixed axis called the fulcrum. When a force is applied at one end of the bar, the point of application moves through the arc corresponding to the angle through which the lever turns, and work is thus expended from the store of energy. The machine then transfers this work to the other end of the lever, where the resistance is moved through a certain distance, and work is thereby performed upon it. The obstacle, or the resistance to be overcome, is frequently in the form of a weight to be lifted, and will be so considered in the treatment of the lever and of the other forms of simple machines. The force applied to the lever, and the resistance, or weight, to be overcome, may be regarded as acting in a direction at right angles to the length of the bar.

Thus in Fig. 69, F and W represent the force and the weight respectively, operating at the two ends of the lever AB, which is free to turn about the fulcrum O. If the system is in equilibrium, a slight rotation about this point will cause F to move in its own direction along the arc Aa, thus performing the work

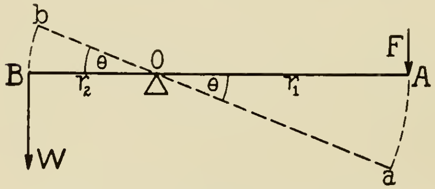

FIG. 69. The principle of the lever.

$$F \cdot Aa;$$

at the same time W moves in its own direction along the arc Bb, and an amount of work equal to

$$W \cdot Bb$$

is performed upon it.

Except for the loss due to friction, which in the lever is not great, the work done upon the weight is exactly equal to that performed by the force, and

$$F \cdot Aa = W \cdot Bb.$$

But (p. 48),

$$Aa = r_1\theta,$$

and

$$Bb = r_2\theta,$$

where r_1 and r_2 are the distances from the fulcrum of the force and the weight respectively, and θ is the angle through which the lever has turned during the rotation. Accordingly,

$$F \cdot r_1\theta = W \cdot r_2\theta,$$

or

$$Fr_1 = Wr_2,$$

that is, the moment of the force about the fulcrum is equal to the moment of the weight about this same point. The last equation may be written

$$\frac{F}{W} = \frac{r_2}{r}.$$

or in words,

$$\frac{\text{Force}}{\text{Weight}} = \frac{\text{Distance of weight from fulcrum}}{\text{Distance of force from fulcrum}},$$

from which we see that the force applied and the weight lifted are inversely proportional to their respective distances from the fulcrum.

The ratio of the resistance overcome to the force applied in any machine is known as the mechanical advantage of the machine. Therefore

$$\frac{W}{F} = \text{mechanical advantage.}$$

118. The Three Classes of Levers. According to the location of the force and of the resistance with respect to the fulcrum, levers may be divided into three groups, known as levers of the first, second and third class respectively.

In levers of the first class, Fig. 70, the force and the weight are applied on opposite sides of the fulcrum, the weight being nearer the fulcrum than the force. In levers of the second class, Fig. 71, the fulcrum is at one end of the bar, the force is applied at the other, while the weight is placed between them. In levers of the third class, Fig. 72, the fulcrum is at one end of the bar, the weight at the other, and the force is applied at an intermediate point. Since in levers of the first and second classes the lever arm of the force is longer than that of the weight, a given force can lift a weight greater than itself, though with diminished speed; but in levers of the third class, where the lever arm of the force is shorter than that of the weight, the force applied must be greater than the weight, and the latter moves faster than does the point of application of the force.

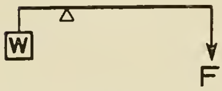

Fig. 70. Lever of the first class.

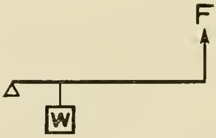

Fig. 71. Lever of the second class.

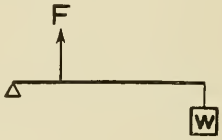

Fig. 72. Lever of the third class.

Levers of the third class are employed when great force is available and rapidity of motion is desired. The fore arm is an illustration of a lever of this class, the elbow representing the fulcrum, the attachment of the muscles of the upper arm being the point of application of the force, while the weight is the object lifted by the hand. The fingers likewise operate upon this principle. It is to the swiftness of movement which could be attained in no other way that the dexterity of the human hand is largely due.

119. The Wheel and Axle. The wheel and axle consists of a relatively small cylinder, rigidly connected with another cylinder, or wheel, of much larger diameter. The weight W to be lifted, Fig. 73, is applied to a rope wrapped about the smaller cylinder, or the axle, while the force F that is to overcome, or raise, this weight is applied to the circumference of the larger wheel, frequently by a cord wrapped in the opposite direction about it. The radii of the wheel and the axle may be taken as r_1 and r_2 respectively. If the system is in equilibrium and is given one rotation in the direction in which the force acts, the force will move through a distance in its own direction equal to the circumference of the wheel, thus performing an amount of work equal to

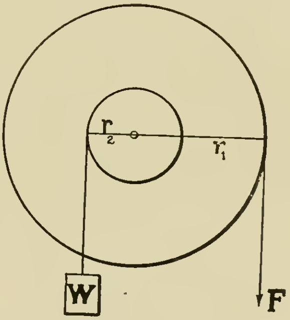

FIG. 73. The wheel and axle.

$$F \cdot 2\pi r_1,$$

while the weight will be lifted through a distance equal to the circumference of the axle, and an amount of work will be done upon it equal to

$$W \cdot 2\pi r_2.$$

With the exception of the energy expended in overcoming friction, which in this machine is larger than in the lever, these two amounts of work are equal, since work can be neither gained nor lost. Therefore,

$$F \cdot 2\pi r_1 = W \cdot 2\pi r_2,$$

or

$$\frac{F}{W}=\frac{r_2}{r_1}.$$

That is

Force : weight = radius of axle : radius of wheel,

or, the force applied and the weight lifted are inversely proportional to the radii of the wheel and the axle respectively.

120. The Inclined Plane. By the use of the inclined plane, a body or weight is to be lifted to a certain height through the action of a force that moves the weight parallel to the face of the plane.

The weight W of the body, acting vertically downward, Fig. 74, may be resolved into two components, one $W \sin \phi$, acting down the face of the plane, and the other $W \cos \phi$, perpendicular to this face, where ϕ is the angle of the plane. If the body is to remain in equilibrium on the plane, a force F, equal to the force $W \sin \phi$ but opposite in direction, must act up the face of the plane. If the body moves up the plane, the force F must act through a distance equal to the length l of the face of the plane, thus performing an amount of work equal to

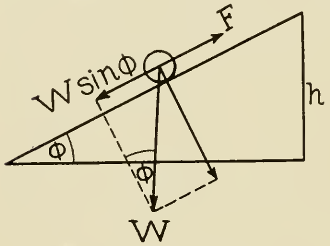

Fig. 74. Forces on the inclined plane.

$$Fl.$$

The resistance, or the weight, which in this case is $W \sin \phi$, has also been moved through the length l of the face of the plane, and an amount of work has been done upon it equal to

$$Wl \sin \phi.$$

Inasmuch as these amounts of work must necessarily be equal,

$$Fl = Wl \sin \phi.$$

Since

$$\sin \phi = \frac{h}{l},$$

where h is the vertical height of the plane, this expression becomes

$$Fl = Wh,$$

or

$$\frac{F}{W} = \frac{h}{l}.$$

Therefore, when the force is applied parallel to the face of the inclined plane.

Force : weight = height of plane : length of plane.

121. The Screw. The screw consists of an inclined plane wound around a cylinder, forming a spiral line known as the thread of the screw, Fig. 75. The thread makes an angle ϕ with the base of the screw equal to the angle of the inclined plane. The distance between two successive turns of the same spiral, measured in the direction of the axis of the screw, is known as the pitch. Therefore, the pitch of the screw corresponds to the height of the inclined plane, the circumference of the cylinder being equivalent to the base. The screw is designed to work in a nut, that is, in a hollow cylinder having threads upon its inner surface, fitting into the spaces between the threads of the screw. One revolution of the screw raises the weight W an amount h equal to the pitch of the screw, while the force F, if this is applied tangentially, moves through

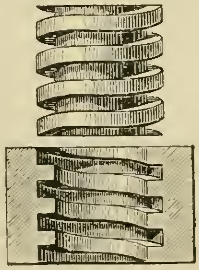

FIG. 75. The screw and its nut.

a distance equal to the circumference of the screw. Accordingly, if r is the radius of the screw,

$$F \cdot 2 \pi r = W \cdot h,$$

or

$$\frac{F}{W} = \frac{h}{2 \pi r}.$$

That is,

Force : weight = pitch of screw : circumference of screw.

If the force is not applied tangentially to the circumference of the screw, but at right angles to a radial arm of length l, the distance described by the force during one revolution is $2\pi l$ and

$$\frac{F}{W} = \frac{h}{2 \pi l},$$

or,

Force : weight = pitch of screw : circumference of circle described by force.

Since the friction between the screw and the nut is considerable, a force greater than that here indicated must be applied.

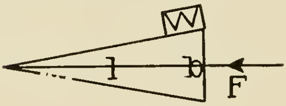

FIG. 76. The wedge.

122. The Wedge. The wedge consists of two inclined planes placed base to base, Fig. 76, the two sloping sides, or faces, meeting at the vertex, or angle, of the wedge, and the side opposite the angle forming the back. The length of the wedge is the perpendicular distance between the angle and the back. If the force is applied parallel to the length of the wedge, that is perpendicular to the back, while the weight to be overcome is moved at right angles to this direction, the force, in moving a distance equal to the length l of the wedge, will perform an amount of work equal to

$$Fl,$$

while the weight W, in being moved a height equal to the back b of the wedge, will have an amount of work done upon it equal to

$$Wb.$$

Accordingly, since these quantities of work are equal,

$$Fl = Wb,$$

or

$$\frac{F}{W} = \frac{b}{l}.$$

Therefore, when the wedge is thrust under an object to lift it, neglecting the loss due to friction, which in this instance is very great, we have the relation

Force : weight = back of wedge : length of wedge.

Because the energy consumed in overcoming friction is enormously great in the use of the wedge, the above theoretical relation between the force applied and the resistance overcome is far from being even approximately correct.

123. The Pulley. The pulley is a device for transmitting energy by means of a cord or rope, and consists in its essential features of a grooved wheel carrying the cord to which the force is applied. If the pulley is rigidly supported in such a way that while it is free to rotate it is incapable of other motion, Fig. 77, its only purpose is to change the direction of the

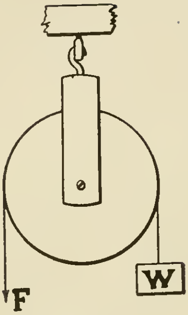

Fig. 77. The fixed pulley.

force, since the tension in one part of a cord can be no greater than in another. In a pulley so supported, that is, in a so-called fixed pulley, the force is applied to one end of the rope, while the weight to be lifted is applied directly to the other. When the force moves through a certain distance, the weight moves through an equal distance, and the force is then equal to the weight, as, for instance, in the pulley used in drawing water from a well.

When, however, the pulley is not rigidly fastened, but is attached to the weight to be lifted, Fig. 78, the cord passes under the pulley, one end being free for the application of the force, the other being fastened to a support. In this arrangement, which is known as the movable pulley, the weight is carried partly by the cord attached to the support, and partly by the cord to which the force is applied, and each of these cords carries half the weight, since the tension in each must be the same. When the force moves through a certain distance, each of these cords is shortened by half that distance, and the weight is therefore lifted by half the distance through which the force moves. Since the product of the force and the distance through which the force moves must equal the product of the

t ; the
ce thru
the weight
s, it follows

If a continuous cord passes successively about a number of lows that in the single movable pulley, the weight lifted is twice as great as the force applied.

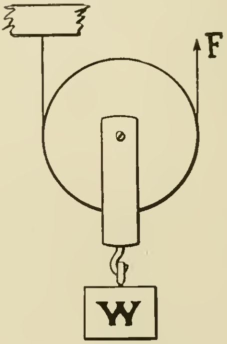

Fig. 78. The single movable pulley.

If a continuous cord passes successively about a number of pulleys some of which are fixed while others are movable, the weight to be lifted is divided equally between the cords that support it. Thus, in Fig. 79, the cord passes around two fixed and two movable pulleys; and as the number of cords that here support the weight is four, each cord carries one quarter of the weight. Since the tension in all parts of the cord must be the same, it follows that a force equal to one quarter of the weight will hold the system in equilibrium. The distance, however, through which the force moves when a displacement occurs is also divided equally among the four cords, each one being shortened by one quarter of this amount, and the weight being lifted by the same amount. The weight, therefore, is lifted only one quarter of the distance through which the force moves, although it is four times as great as the force in magnitude. In general, the weight that can be lifted by a system of pulleys is equal to the force multiplied by the number of cords supporting the weight. Moreover, if one end of the cord is fastened to the fixed pulleys of the system, as in the figure, the weight lifted is $2n$ times the force applied, where n is the number of movable pulleys, while the distance through which the weight is lifted is only $\frac{1}{2n}$ of the distance through which the force moves.

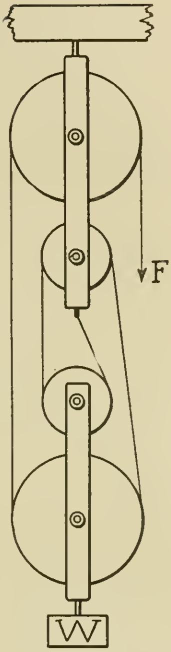

FIG. 79. The compound pulley.

In all forms of the pulley, as in all other machines, the law holds true that, except for the energy consumed in friction, the work done by the force is equal to the work done upon the weight, that is,

$$\text{Force} \times \text{distance through which the force moves}$$
$$= \text{weight} \times \text{distance through which the weight moves},$$

or

$$\frac{\text{Force}}{\text{Weight}} = \frac{\text{distance through which the weight moves}}{\text{distance through which the force moves}}.$$

The force and the weight, therefore, are inversely proportional to the distance through which each moves.

124. Friction. It is a matter of common experience that motions upon the earth soon cease, and that moving bodies ultimately come to rest, even though there may be no visible forces acting to stop them. This is true because there is always a force, known as the force of friction, acting to oppose motion whenever one body moves over another. It is to be observed that the force of friction does not exist until motion occurs or is about to occur, and this force then acts in such a way as to oppose the motion and bring the body to rest. It is evident, therefore, that the cause of friction has its origin in the motion of one body over another, and not in the fact that the two are in contact with each other. Moreover, it is found that the friction between two surfaces is greater when these surfaces are rough than when they are smooth, although only to a certain extent can the friction be eliminated by polishing the surfaces of contact.

125. The Cause of Friction. A partial explanation of this phenomenon undoubtedly lies in the fact that even polished surfaces are not perfectly smooth, since each contains minute protuberances which project into the spaces between similar inequalities in the other, Fig. 80. Therefore, when one surface is moved over the other, these little protuberances are torn down and broken off, thus requiring the expenditure of energy. This theory is supported by the fact that surfaces become smoother because of friction, and also by the fact that lubricating substances, such as oil, black lead, etc., diminish the friction, since these are thought to fill up the little inequalities in the surfaces, thus preventing the interlocking of the protuberances.

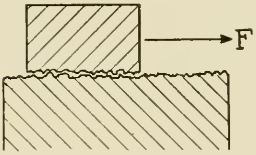

FIG. 80. Friction due to roughness of surface.

A further explanation is found in the possible adhesion between the particles of the two surfaces. The pressure probably brings the two surfaces so near together that the molecules of one surface attract those of the other, and this molecular attraction must be overcome when one surface is moved over the other. Both of these theories are confirmed by the fact that the friction at rest is a little greater than the friction in motion, that is, the resistance offered by friction to motion about to begin is a little greater than that offered to motion already taking place. This is explained by the supposition that when the surfaces are at rest, the particles have time to fit into one another, and thus bring the molecules of one surface more completely within the sphere of attraction of the molecules of the other.

126. Sliding Friction and Rolling Friction. Friction of the kind here described, produced by the sliding of one body over another, is known as sliding friction. Another form, however, known as rolling friction, is found in the resistance to motion that occurs when one body rolls over another. This form is thought to exist because the rolling body, Fig. 81, must crush the little protuberances and thereby expend energy as it passes over them.

Indeed, it is because rolling friction is less than sliding friction that an engine is able to draw a train of cars many times heavier than itself. If the brakes on the cars are set, so that the locomotive can not start the train, the wheels of the engine will turn uselessly around, that is, they will slide upon the track, and the resistance of the train is greater than this sliding friction. The maximum force which the engine can exert is thus seen to be equal to the sliding friction between its wheels and the rails. If the brakes are released, however, so that the wheels of the cars are free to roll, the resistance, offered by the train is only that of its rolling friction; and if this is less than the sliding friction between the wheels of the locomotive and the rails, the train will move. The engine, therefore, can

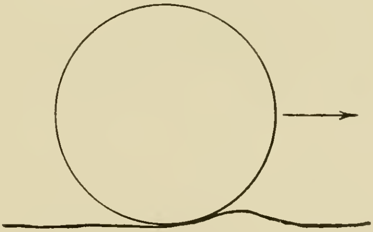

Fig. 81. The cause of rolling friction.

draw the train so long as the combined rolling friction of all the cars is less than the sliding friction between the wheels of the engine and the track.

127. The Coefficient of Friction. The friction between two surfaces is found to depend upon *the kind of material* of which these surfaces are made, and must be determined for each two different substances. In this measurement, a smooth table is usually made from one of the substances and a smooth plate, or block, from the other, Fig. 82, and the force necessary to draw the block over the table is then determined, this force being known as *the force of friction.* When the block rests upon the table, the force, or pressure, perpendicular to the surface of separation between the two is known as the normal pressure.

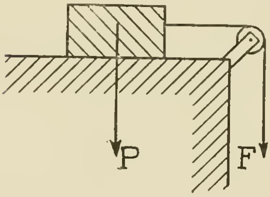

Fig. 82. Determining the coefficient of friction.

By experiment it is found that *the force of friction is proportional to the normal pressure;* that is, if the pressure between the block and the table is doubled, the force required to draw the block along is also doubled. There is, therefore, a constant ratio between the force of friction and the normal pressure, and this ratio, which is constant for any two substances, is known as *the coefficient of friction.* Accordingly if F denotes the force of friction, and P the normal pressure, the coefficient of friction k is always such that

$$k = \frac{F}{P}.$$

Since the force of friction depends only upon the nature of the surfaces in contact and upon the normal pressure, it is independent of the extent of the surfaces. Therefore, if a certain force of friction is required to draw along a given block when this is lying flat upon the table, the same force of friction will draw the same block when this is turned upon edge. It is also found by experiment that the coefficient of friction increases with the velocity, although it does so only to a relatively small extent, and is usually not considered.

128. Coefficient of Friction by the Inclined Plane. Another method of determining the coefficient of sliding friction is to place the block upon an inclined plane, and to increase the angle of the plane until the block will slowly slide down. The angle which the plane makes with the horizontal when sliding is just about to begin is known as the angle of friction, or the angle of repose. When this condition has been attained, the coefficient of friction is numerically equal to the tangent of the angle of friction, as may be seen from the following consideration:

Let *ABC*, Fig. 83, represent the inclined plane, and *D* the block placed upon it, where the angle ϕ of the plane is such that sliding is just about to begin. The weight *W* of the body may be resolved into two components, one *F* parallel to the face of the plane, and the other *P* perpendicular to this face. The force *F* is the force causing the block to slide down the plane, that is, it is the force of friction, while the component *P* is the normal pressure between the two surfaces in contact. The ratio of these two components, therefore, is the ratio between the force of friction and the normal pressure, and this is the coefficient of friction *k* for these two surfaces.

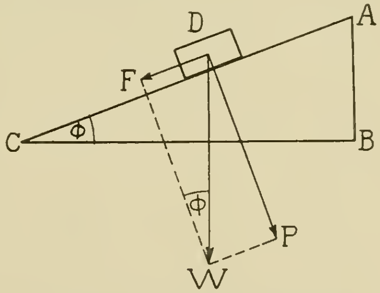

Fig. 83. The coefficient of friction is the tangent of the angle of friction.

Since the angle *DWF* is equal to the angle of the plane, their sides being perpendicular,

$$\frac{F}{W} = \sin\phi, \text{ or } F = W\sin\phi,$$

and

$$\frac{P}{W} = \cos\phi, \text{ or } P = W\cos\phi.$$

Therefore,

$$k = \frac{F}{P} = \frac{W \sin \phi}{W \cos \phi} = \tan \phi.$$

The coefficient of friction is thus seen to be numerically equal to the tangent of the angle of friction.

129. Energy of a Rotating Body—Moment of Inertia. When a body of mass m moves in a straight line with the velocity v, its kinetic energy (p. 121) is expressed by

$$\frac{m v^2}{2},$$

and is independent of the location of the mass, as well as of its distribution. When, however, a rigid body rotates, the various particles of the body have different velocities, and therefore different amounts of kinetic energy, depending upon their distances from the axis of rotation. Accordingly, the kinetic energy of a rotating body depends not only upon the mass and the angular velocity with which it rotates, but upon the distance of the moving mass from the axis of rotation. Since particles far from the axis have a higher velocity, and therefore a greater kinetic energy, than similar particles near the center, it follows that the kinetic energy of a rotating body may be expressed by determining the kinetic energy of each of its particles, and taking their sum.

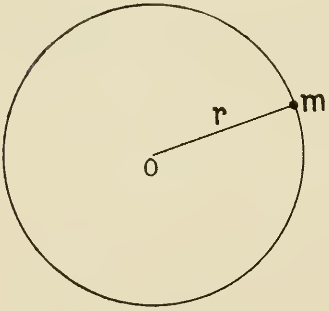

Fig. 84. Kinetic energy of a single rotating particle.

We may assume at first, for simplicity, that the rotating body consists of a single particle of mass m, Fig. 84, at the distance r centimeters from the point O, about which it rotates. Its kinetic energy is then

$$\frac{mv^2}{2},$$

where v is the velocity of the particle measured along the circumference of the circle, or its speed. If the particle makes one revolution in T seconds, its velocity is (p. 51)

$$v = \frac{2\pi r}{T},$$

and the expression for its kinetic energy then becomes

$$\frac{m\,4\,\pi^2 r^2}{2\,T^2}$$

or

$$\frac{1}{2}\,\frac{4\,\pi^2}{T^2}\,.\,m\,r^2.$$

The coefficient of mr^2, that is,

$$\frac{1}{2}\,\frac{4\,\pi^2}{T^2},$$

is constant, and so does not depend upon the mass or upon the distance of the latter from the axis. It is thus seen that the kinetic energy of a rotating body depends not only upon the mass of the body but upon the square of its distance from the axis of rotation.

If another particle m', rotates in the same time as the first, and at the distance r' from the axis, its kinetic energy is

$$\frac{1}{2}\,\frac{4\,\pi^2}{T^2}\,m\ r'^2.$$

The total kinetic energy of these two particles is then

$$\frac{1}{2}\frac{4\pi^2}{T^2}m\,r^2 + \frac{1}{2}\frac{4\pi^2}{T^2}m'\,r'^2 = \frac{1}{2}\frac{4\pi^2}{T^2}(m\,r^2 + m'\,r'^2).$$

As nothing will be changed if these two particles are rigidly connected together, the kinetic energy of a rigid body consisting of two particles rotating about the same axis, is

$$\frac{1}{2}\frac{4\pi^2}{T^2}(m\,r^2 + m'\,r'^2).$$

If we now consider an extended body, consisting of a great many (n) particles, of different masses and situated at different distances from the axis of rotation, we may in the same way calculate the kinetic energy for each particle. The total kinetic energy of the entire rotating body, that is, the sum of the kinetic energies of its particles, is

$$\frac{1}{2}\;\frac{4\pi^2}{T^2}(m\,r^2 + m'\,r'^2 + m''\,r''^2 + \ldots + m^n\,r^{n2}).$$

The expression within the parenthesis may be written

$$\Sigma\,mr^2,$$

where Σ denotes "the sum of." Therefore, the total kinetic energy of a rotating body is

$$\frac{1}{2}\;\frac{4\pi^2}{T^2}\,\Sigma\,mr^2.$$

That is, the mass of each particle is to be multiplied by the square of its distance from the axis of rotation, and the sum of these products taken. The sum of these products, multiplied by the constant

$$\frac{1}{2}\;\frac{4\pi^2}{T^2}$$

expresses the total kinetic energy of the rotating body. This sum,

$$\Sigma\, mr^2,$$

is known as the moment of inertia.

In practice, however, a summation such as this cannot be made, since it is impossible to ascertain the mass, and therefore the individual moment of inertia, of each particle. Nevertheless the equivalent of this may be accomplished by determining the moment of inertia that would exist if the entire mass were concentrated at a single point, and this point were situated at such a distance k from the axis that the kinetic energy would remain unchanged. In this case, the entire mass M of the body, multiplied by the square of this distance k, would equal the moment of inertia of the body, that is,

$$Mk^2 = \Sigma\, mr^2,$$

where the distance k is known as the radius of gyration. Accordingly, the radius of gyration is the distance from the axis of rotation such that, if the entire mass of the rotating body were concentrated in a point at this distance, the moment of inertia, and therefore, also, the kinetic energy, would remain unchanged.

It remains to discuss the coefficient of Σmr^2. In the above expression for kinetic energy, we have seen, (p. 51), that $\frac{2\,\pi}{T}$ represents the angular velocity ω of a rotating body, and the fraction $\frac{4\,\pi^2}{T^2}$ is simply the square of this quantity. The coefficient of Σmr^2 in the above expression for kinetic energy is, accordingly,

$$\frac{1}{2}\,\frac{4\,\pi^2}{T^2},\ \text{or}\ \frac{1}{2}\,\omega^2.$$

The expression for the kinetic energy of a rotating body may, therefore, be written

$$\frac{1}{2} \Sigma mr^2 \omega^2,$$

or

$$\frac{1}{2} I \omega^2,$$

where I is the moment of inertia.

A clearer conception of the full meaning of the moment of inertia may be obtained by comparing the expressions for the kinetic energy of translation and the kinetic energy of rotation, and observing the significant analogy between the two:

$$\text{Kinetic energy of translation} = \frac{1}{2} m v^2,$$

$$\text{Kinetic energy of rotation} = \frac{1}{2} I \omega^2.$$

In these expressions, which are identical in form, it is seen that the moment of inertia in a motion of rotation corresponds to the mass, that is, to simple inertia, in a motion of translation, while the angular velocity in rotation is a vector quantity analogous to linear velocity in translation. Since the energy of a rotating body depends upon its moment of inertia, the work done by a force in setting a body in rotation must likewise depend upon its moment of inertia. Accordingly, when a force acts upon a mass of matter to impart rotation, this force must overcome the moment of inertia of the matter, that is, its rotational inertia, just as a force which acts upon a mass of matter to produce translation must overcome its linear inertia. From this we see that moment of inertia is the resistance that matter offers to being set in rotation, and it is also the resistance that matter offers to a change of rotation, when this is once accomplished. It is, therefore, the resistance that matter offers, not to a rotation, but to a *change* of rotation;

or, in other words, it is the persistence with which matter maintains unchanged its existing condition as regards rotation. But we have seen (p. 99) that simple inertia is the tendency of matter to maintain unchanged its existing conditions as regards translation. Hence it is, that moment of inertia bears the same relation to rotation that simple inertia bears to translation.

It is because the kinetic energy of a rotating body is proportional to its moment of inertia, that is, proportional not only to its mass, but to the square of the distance of the latter from the axis of rotation, that fly-wheels of engines are made with very massive rims. Great energy is required to set these wheels in rotation; but, once in motion, their kinetic energy is so great, because of their great moment of inertia, that they continue to rotate, even when an extra load is suddenly thrown upon the engine, and so a uniform speed of the machinery is maintained.

To here

CHAPTER IX.

GRAVITATION AND GRAVITY.

130. The Attraction of Gravitation. The great law of the attraction of gravitation, as announced in 1685 by Sir Isaac Newton, is as follows:—Every particle of matter in the universe attracts every other particle of matter with a force which acts in the direction of the straight line joining their centers, and varies directly as the product of the masses and inversely as the square of the distance between them. Thus, if m and m' represent the masses of two bodies and r is the distance between their centers, the force of attraction between them may be expressed by the relation

$$F : \frac{m\,m'}{r^2},$$

or

$$F = k\,\frac{m\,m'}{r^2},$$

where k is the factor of proportionality, and is known as the constant of gravitation.

The reason why masses of matter attract one another has been a mystery to men of all ages, and even now it is not at all understood. It may, perhaps, best be explained by assuming that there is a stress in the medium (the ether) between the attracting bodies. The nature of this attraction, therefore, is not known, but the law according to which it takes place is the one stated above. If one of these masses is the earth, and the other a body upon the earth, the force of attraction between the two is known as the weight of the body. Weight therefore, is a true force, and may be measured in dynes. The term *gravitation* is the general term applied to this force of attraction throughout the universe, while the word *gravity* is applied to the special case of gravitation when the attraction exists between the earth and objects upon it.

131. The Cavendish Experiment. The constant of gravitation k is the force with which 1 gram of matter attracts another gram of matter when the two are at the distance of 1 centimeter from each other. The determination of the value of this constant was made in the years 1797–8 by Henry Cavendish, and furnished the first experimental demonstration of the truth of Newton's law of the attraction of gravitation between masses of matter.

In this remarkable investigation, Cavendish devised and employed the arrangement which is still used for this and for similar purposes, known as the torsion balance. This consisted of a horizontal bar carrying a small mass m at each end, and suspended at its center by an extremely fine wire, so that the bar would be deflected under the action of exceedingly small forces applied to the masses m. From a detailed study of the apparatus, the force required to twist the wire through a certain small angle was accurately known. When large masses of lead M were brought very near to the masses m, and in the positions shown in Fig. 85, the attraction of gravitation between these masses and the smaller masses m gave rise to a couple, producing a small rotation of the bar in the positive direction and a slight twisting of the suspending wire. The large masses of lead were then removed to similar positions on the opposite side of the bar, Fig. 86, and the force of attraction which then took place between the masses produced a rotation of the bar and a twisting of the wire in the opposite sense.

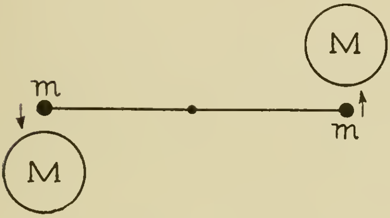

FIG. 85. Cavendish's torsion balance.

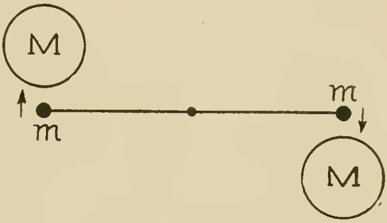

FIG. 86. The torsion balance with the lead masses transposed.

From the amount of rotation thus produced, and from the knowledge of the force required to twist the wire through a given angle, it was possible to compute in dynes the force of attraction between the masses M and m. When this force was substituted for F in Newton's expression

$$F = k\,\frac{m\,m'}{r^2},$$

and when the masses of the large leaden spheres and those of the smaller balls at the ends of the bar were substituted for m and m' respectively in this equation, r being the distance between their centers, the value of k was found to be approximately,

$$\frac{1}{15{,}000{,}000}\ \text{dyne}.$$

132. The Mass and Density of the Earth. Although the determination of k was in itself a most remarkable work, Cavendish was able to employ the value of this constant in a calculation of still greater interest, which led to the computation of the mass of the earth. This he accomplished by once more making use of Newton's equation, and considering the earth and a mass of matter on the earth as the two attracting bodies. We may follow this computation without difficulty, but for the sake of simplicity we shall make use of the modern notation.

If m in the above equation represents the unknown mass of the earth and m' a mass of 1 gram on its surface, r is then the distance from the surface of the earth to the center, that is, the radius of the earth, or 637×10^6 centimeters. Since the force F, with which the earth attracts a mass of 1 gram at its surface, is 981 dynes, and since the value of k is given by the Cavendish experiment, all quantities except m in this equation are known, and the latter becomes

$$981 = \frac{1}{15,000,000} \times \frac{m \times 1}{(637 \times 10^6)^2},$$

from which

$$m = 614 \times 10^{25} \text{ grams.}$$

The mass of the earth, accordingly, is 614×10^{25} grams; and Cavendish, who thus determined this mass with the aid of the torsion balance, is often referred to as the first who was able to "weight the earth".

One further deduction of the greatest importance is a direct result of the work of Cavendish. The radius, and therefore the volume, of the earth being known, and its mass having been thus determined, its average density, that is, the ratio of its mass to its volume (p. 77), may be at once computed, and is found to be 5.527. This means that the earth contains 5.527 times as much matter as if it consisted entirely of pure water.

133. Expression for the Weight of a Body. When a body falls freely, its weight W, or the attraction of the earth upon it, gives to its mass m the acceleration g. Since weight, then, is a true force, we may obtain its magnitude in dynes by substituting these values for the force and the acceleration in the general expression for force, namely,

$$F = m\,a,$$

and obtain

$$W = m\,g.$$

This is the corresponding expression for the weight of a body in terms of its mass and the acceleration of gravity, and indicates that a body of mass m, in falling freely under the attraction of gravity, that is, under the action of its own weight, acquires an acceleration of g centimeters per second in a second. The same equation shows, likewise, that *the weight of a body is directly proportional to its mass;* for if the mass m were doubled, the weight W would also be doubled, so that the weight of a 2-kilogram mass is twice as great as that of a 1-kilogram mass.

134. A Heavy and a Light Body Fall in Equal Times. We have seen, however, (p. 80), that when forces are proportional to the masses upon which they act they produce the same acceleration. The force of gravity acting upon the 2-kilogram mass, or its weight, is twice as great as that acting upon the 1-kilogram mass; the mass of the former is also twice as great as that of the latter, and therefore, the acceleration which is imparted to each is the same in each case, or g. Accordingly, the 2-kilogram mass, in falling, will acquire the same acceleration as the 1-kilogram mass, that is, both will fall in the same time.

The explanation of this phenomenon lies in the fact that the inertia of a body, or its resistance to change of motion, is proportional tó its mass (p. 101). The two-fold weight, therefore, acting upon the two-fold mass to accelerate it, encounters twice as much resistance, or inertia, as the single weight acting upon the single mass; but as the force acting upon the larger mass is also twice as great as that acting upon the smaller, the acceleration given to the former is exactly the same as that imparted to the latter. A heavy and a light body, therefore, will fall in equal times, as was first demonstrated by Galileo by dropping together two very unequal masses from the top of the Leaning Tower of Pisa. The two were seen to start at the same time, to fall together, and to strike the ground at the same instant. Thus was demonstrated the truth, which the world even then was slow to believe, that the acceleration of gravity is independent of mass, or that a small mass falls in the same time as a large one.

It may be argued that this cannot be so, inasmuch as it is contrary to daily experience; for it is certainly true that a light body, such as a piece of paper, falls more slowly than a heavier one, such as a coin. This, however, is because of the resistance of the air, which operates as a force tending to decrease the motion of each body. The effect of this impeding force is relatively greater upon the lighter body than upon the heavier one; but if the two are enclosed in a tube, from which the air is exhausted, the force of gravity is then the only force acting, and they will be seen to fall in the same time.

A simpler experiment illustrating the same principle consists in placing a sheet of paper upon a flat metallic plate, and dropping the two together. The plate, as it falls, displaces the air before the paper, and so frees the latter from this retarding effect. The two, therefore, fall together, and the paper is still in contact with the plate as it strikes the ground.

135. Mass and Weight. The distinction between mass and weight is now apparent. The mass of a body is the quantity of matter in the body, and remains the same so long as matter is neither added to it nor taken from it. The weight of a body, however, is the measure of the attraction of the earth upon it, and varies with its distance from the center of the earth. The mass of a body, therefore, is a property of that body, and is inherent in it, while its weight cannot exist unless another body is present to attract it.

136. Varation of Gravity at Different Points of the Earth's Surface. If the earth were a perfect sphere, the weight of a body, except for a small effect due to centrifugal force, would be the same at every point upon the surface of the globe. Inasmuch, however, as the latter is spheroidal in form, a body at the poles is nearer the center of the earth than is a body at the equator. Accordingly, Fig. 87, if a body of mass m' were carried from the equator to the pole, m being the mass of the earth, m and m' in the expression

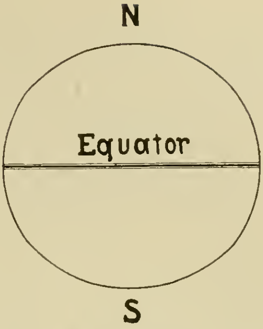

FIG. 87. Weight increases from the equator to the poles.

$$F = \frac{m\,m'}{r^2}$$

would remain the same, but r would vary with the latitude, being less at the poles than at the equator. At the poles, this shorter distance from the center of the earth means an increased attraction and, therefore, an increased weight. The weight of a body, accordingly, is not constant over the surface of the earth, but increases as the body is carried from the equator to the poles.

137. The Attraction of a Hollow Shell of Gravitating Matter Upon a Particle Placed Within. As a special case of the attraction of gravity it is interesting to determine the weight of a body at a point in the interior of the earth. Before considering this problem, it is necessary to demonstrate the following proposition: A particle of matter placed within a hollow shell of gravitating matter experiences no force of attraction due to the shell, wherever within the latter the particle may be placed.

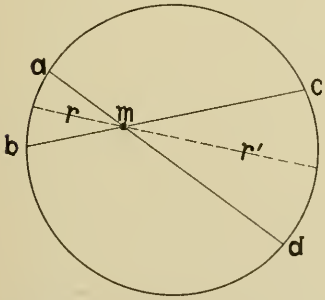

FIG. 88. A hollow shell exerts no resultant attraction upon a particle placed within.

Let the circle Fig. 88, represent the hollow shell of gravitating matter, and m the particle of matter within. m may be made the vertex of two small similar cones that cut out similar areas upon the surface of the sphere, represented in section by the lines ab and cd. The distance of m from these portions of the sphere and r and r' respectively. The attraction of the mass ab upon m is

$$\frac{ab \,.\, m}{r^2},$$

acting toward the left; and the attraction of the mass cd upon m is

$$\frac{cd \,.\, m}{r'^2},$$

acting toward the right. But by geometry, the bases of similar cones are to each other as the squares of their altitudes. Therefore,

$$\frac{ab}{r^2} = \frac{cd}{r'^2}.$$

If each side of this equation is multiplied by m, we have

$$\frac{ab \,.\, m}{r^2} = \frac{cd \,.\, m}{r'^2}.$$

This, however, is an equation, each member of which is identical with the force of attraction as given above, thereby showing that the attractions of these two portions of the spherical surface upon the particle m are equal and opposite. The whole surface of the sphere may be divided into pairs of similar cones in the same way, for each pair of which the attraction is equal and opposite, and so the particle experiences no resultant attraction in any direction. The proposition, that a hollow shell of gravitating matter exerts no resultant attraction upon a particle placed within, is therefore proved.

138. The Weight of a Body at a Point in the Interior of the Earth. When a particle of matter is situated in the interior of the earth, as, for instance, at the bottom of a deep mine, Fig. 89, it may be conceived as being upon the inside of a hollow shell, the thickness of which is everywhere equal to the depth of the mine, and also upon the outside of a smaller sphere, the radius of which is the distance of this particle from the center of the earth. Inasmuch as the exterior shell, by the preceding demonstration, exerts no resultant attraction upon the particle, the weight of the particle is due solely to the attraction of this smaller sphere. If $\frac{4}{3}\pi r^3 d$ represents the mass of the earth, and r its radius, and $\frac{4}{3}\pi r'^3 d$ the mass of the smaller sphere, and r' its radius, where d is the density, and if w represents the weight of the body m at the surface of the earth, and w' its weight at the interior point, then

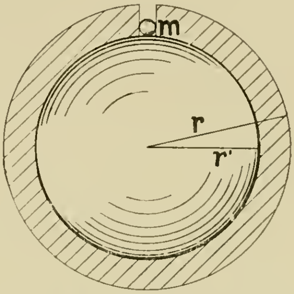

FIG. 89. The weight of a body inside the earth.

$$w : w' = \frac{4}{3}\frac{\pi r^3 d . m}{r^2} : \frac{4}{3}\frac{\pi r'^3 d . m}{r'^2} .$$

or

$$w : w' = r : r'.$$

From this it is seen that at a point in the interior of the earth the weight of a body varies directly as the distance of that point from the center.

139. The Theory of Gravitational Potential. We have seen (p. 120) that when a body has been lifted against the force of gravity work has been done upon it; and because of the work so performed, the body possesses an energy of position equivalent in amount to the work that was expended in bringing it to this height. The elevated position of the body then denotes a store of potential energy in the body. Before this energy could exist, however, and, therefore, before the body was brought to this elevation, the condition at this point in space was such that the body, when it should be brought to this point, would then possess potential energy. From this it follows that, due to the presence of the earth, every point in space in the region about the earth possesses an attribute of such nature that matter, when placed at that point, will possess potential energy. *This condition at a point, in virtue of which matter will possess potential energy when placed at that point, is known as potential.* It is to be observed that the energy is not present until matter is actually there; the potential, however, exists whether matter is there or not, for it is a characteristic of the point, and is entirely independent of matter.

The measure of the potential at a point is the number of ergs of potential energy that the unit mass will have when placed at that point; that is, it is the number of ergs of work that will be required to bring the unit mass from the surface of the earth to that point. If this mass were lifted to a higher point, more work would have to be done upon it, and it would then possess a greater amount of energy. The attribute of a higher point is such as to give the mass greater energy when placed at that point. Therefore, the farther from the surface of the earth the point is taken, the higher is the potential of that point.

140. Equipotential Surfaces. If equal amounts of work are done upon equal masses in lifting them from the surface of the earth, the masses will then possess equal amounts of potential energy. The potentials of the points to which the masses are carried will, accordingly, be equal to one another, that is, these points will then have the same potential. There are, therefore, near the surface of the earth, an infinite number of points such that the potential of any one is the same as that of any other. Through all points that have a common potential it is possible to lay a surface, Fig. 90; and a surface thus connecting all points at the same potential is called an equipotential surface.

Since all points outside this surface have a potential higher than the surface itself, it is possible to find among these exterior points a number that have a common potential among themselves. Through all these points another equipotential surface may be laid, over which the potential is constant and higher in value than on the interior equipotential surface. It is obvious that equipotential surfaces can never touch one another, for if they were to intersect, the point or points where they came together would then have two potentials, which is impossible.

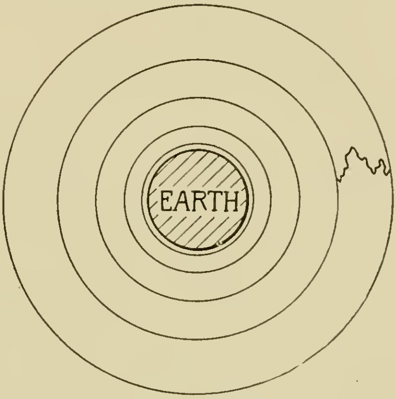

Fig. 90. Equipotential surfaces about the earth.

Equipotential surfaces have unit difference of potential between them when on erg of work is required to carry the unit mass from one surface to the other.

141. The Work Done in Carrying a Body From One Equipotential Surface to Another. Since the energy of a body is the same at one point of an equipotential surface as at any other point on the same surface, no work is performed against the force of gravity in moving a body along an equipotential surface. The direction of the force of gravity, therefore, is always at right angles to the equipotential surfaces. From this it follows that when a body is carried from one equipotential surface to another, the work done is independent of the path, and depends only upon the force of gravity at that point, and upon the perpendicular distance between the two surfaces.

This is so because any motion however irregular may be resolved into two components, one of which is perpendicular to the equipotential surfaces, and the other of which is parallel to these surfaces. Since no work is performed against the force of gravity in moving a body parallel to an equipotential surface, the entire work is done by the force of gravity acting along the component of the path that is perpendicular to these surfaces. The work done, therefore, in carrying a body from one equipotential surface to another is independent of the path along which the body is carried.

142. The Center of Gravity. When a mass of matter is acted upon by gravity, each of its particles is attracted towards the center of the earth. Accordingly, the attraction of the earth upon each particle of a body gives rise to a system of parallel forces distributed throughout the entire mass. Since these forces, moreover, are all parallel and in the same direction, they may be compounded into a single resultant equal to their sum, this resultant being the weight of the body, Fig. 91. The weight of a body, therefore, is the resultant of the system of parallel forces, due to gravity, acting upon the individual particles of which the body is composed. Inasmuch as each of these forces is proportional in magnitude to the mass of the particle upon which it acts, the algebraic sum of the moments of these forces about the center of mass of the body must be zero (p. 91). The point of application of this resultant, therefore, that is, the point of application of the weight of a body, coincides in position with the body's center of mass (p. 96), and is known as the center of gravity. If the entire mass of the body were concentrated at its center of gravity, the resultant effect of the attraction of the earth upon it would remain unchanged in magnitude and direction.

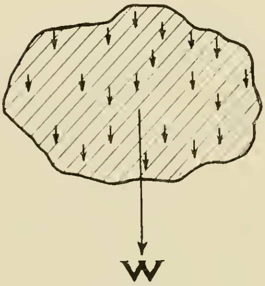

FIG. 91. The center of gravity.

143. Conditions for Equilibrium. The straight line joining the center of gravity and the center of the earth is known as the *line of direction*, and is the line along which the center of gravity would move if the body were allowed to fall freely. Accordingly, if the center of gravity of a body is supported, the entire body is supported; and in order that the center of gravity may be supported, the line of direction must fall within the area of the base. A leaning tower, therefore, is stable, because its line of direction falls within its base, Fig. 92; but if the tower were higher, or if it leaned at an angle so great that the line of direction were to fall without the base, it would topple over.

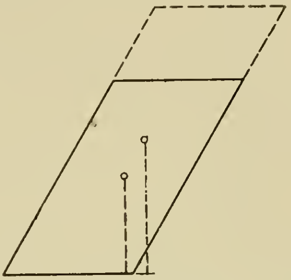

FIG. 92. The stability of a leaning tower.

The falling of a leaning tower, and, indeed, the falling of any body whatever until its center of gravity reaches the lowest possible position, serves to illustrate the principle already discussed, (p. 122), that potential energy tends towards a minimum. It is because the earth attracts a body as if its mass were concentrated at its center of gravity, that the center of gravity always tends to seek the lowest possible point. If, therefore, the body is so supported that any movement tends to raise the center of gravity, Fig. 93, the body is said to be in *stable equilibrium,* since the center of gravity has already reached its lowest point. But if the body is so supported, Fig. 94, that any movement tends to lower the center of gravity, this point has not reached its lowest position, but will do so at the earliest opportunity, and the body is in *unstable equilibrium.* A body is in *indifferent equilibrium* when any movement tends neither to raise nor to lower the center of gravity.

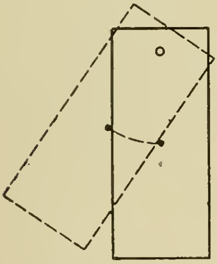

FIG. 93. Stable equilibrium.

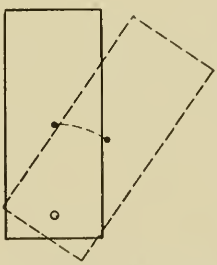

FIG. 94. Unstable equilibrium.

144. Determination of the Center of Gravity. The principle that the center of gravity of a suspended body seeks the lowest possible point, and is found vertically below the point of support, may be employed to determine the center of gravity of an irregularly shaped body. If a body is suspended from any point, such as A, Fig. 95, the center of gravity seeks the lowest possible position, and is found somewhere on the line of direction *AD*. If the body is now suspended from another point, such as *B*, the center of gravity is again vertically below the point of support, and is found somewhere on the line of direction *BE*. Inasmuch therefore, as the center of gravity is found on these two lines of direction, it is found at their point of intersection *C*.

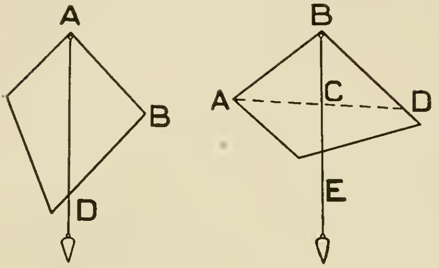

FIG. 95. The determination of the center of gravity.

CHAPTER X.

THE PENDULUM

145. The Simple and the Compound Pendulum—Definitions. A pendulum is any heavy body so suspended as to vibrate freely. Of the various pendulums included under this broad definition, all may be divided into two classes known as the simple, or mathematical, pendulum, and the compound, or physical, pendulum. The simple, or mathematical, pendulum may be defined as a single material particle suspended by a thread without weight; while the compound, or physical, pendulum is any extended body suspended by an axis, so as to vibrate freely about that axis as a center. It is evident that the simple pendulum has no actual existence, for it is impossible to isolate a single material particle, or to suspend such a particle by a thread without weight. All pendulums, therefore, with which we have to deal are really compound.

By suspending a small heavy ball by a very fine, light thread, we may, however, have a pendulum that conforms very closely to the conditions of the simple pendulum. If we consider such a pendulum, the ball, or "bob", will move upon the arc of a circle, the center of which is the point of suspension, and the radius of which is the length of the pendulum. The *length* of the pendulum, therefore, is the distance from the point of suspension to the center of the ball, or bob. The *position of equilibrium* of the pendulum is the central point of its swing, that is, it is the position the pendulum assumes when at rest. The *displacement* of the pendulum at any moment is its distance in either direction from the point of equilibrium, and this distance of course, is constantly changing. The distance from the position of equilibrium to the extremity of the path in either direction is known as the *amplitude;* hence the amplitude is the greatest possible displacement. The *period* of the pendulum is the time required for a complete, or double, vibration, that is, for a full swing forward and back. The time required for the pendulum to swing simply one way is known as the time of a single vibration, or an oscillation.

146. The Vibrations of the Pendulum Take Place with Harmonic Motion. In the simple pendulum, or in the pendulum approaching as closely as possible to the simple pendulum, we see that wherever in its path the heavy bob may be, it is acted upon by a force due to the attraction of the earth, directed vertically downward. This attraction of the earth for the bob, that is, the weight of the bob, is the only force acting on the pendulum; and here, as always,

$$\text{Weight} = mg,$$

where m is the mass of the bob and g the acceleration of gravity. This weight of the pendulum may be resolved into two components, one of which, F, acts in the direction of the motion of the pendulum, while the other, P, acts at right angles to this direction, that is, in the direction of the length of the thread. The former component, F, acting at each point in the direction of motion, is alone effective in producing the motion of the pendulum, while the latter component, P, simply gives rise to the tension in the thread, and in no way affects the motion.

[illegible]

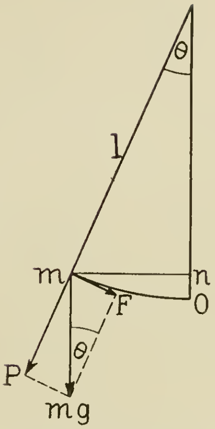

Fig. 96. The force causing the pendulum to vibrate is a component of the weight.

Since the force F, which produces the motion, is always a component of the weight mg, it is clear, Fig. 96, that

$$\frac{F}{mg} = \sin\theta,$$

or

$$F = mg\sin\theta,$$

where θ is the angle through which the pendulum has been displaced from its position of equilibrium. If the angle θ through which the pendulum vibrates is very small, that is, if it does not exceed 3° or 4°, the angle θ may be used instead of $\sin\theta$,* and we may write

$$F = mg\theta,$$

from which it is seen that F is directly proportional to θ. Since F is proportional to the angle θ, it is also directly proportional to $l\theta$, that is, to the arc mO. The arc $l\theta$, or mO, however, is the displacement of the particle m from its position of equilibrium O.

---

* $$\frac{mn}{l} = \sin\theta, \text{ or } mn = l\sin\theta;$$

also

$$\text{arc } mO = l\theta. \qquad \text{(p. 48)}$$

When θ is very small, the arc mO is very nearly equal to the perpendicular mn, and these two become more and more nearly equal as θ grows smaller. When θ is so small that without serious error we may take

$$\text{perpendicular } mn = \text{arc } mO,$$

then

$$l\sin\theta = l\theta,$$

and

$$\sin\theta = \theta$$

We see, therefore, that the force causing the pendulum to vibrate is always directed towards the position of equilibrium, and is proportional to the displacement. But this is the law of the force that produces harmonic motion. From this it follows that when the amplitude is small, the movement of the pendulum takes place with harmonic motion and the vibrations are isochronous.

147. Conservation of Energy Illustrated by the Pendulum. As has already been pointed out (p. 122), the swinging of the pendulum illustrates most beautifully the conservation of energy in a closed system, and also the transformation of energy that is continually taking place from potential to kinetic, and from kinetic again to potential. As the pendulum is moved from its position of equilibrium, and carried to the extremity of its path, work must be done against gravity in lifting the bob, and at the extremity of its path its energy is entirely potential. As it swings towards its position of equilibrium, it loses in potential energy and gains in kinetic until this position is reached, when its energy has become wholly kinetic. In virtue of the store of kinetic energy it then possesses, the pendulum will lift itself an equal distance upon the farther side, and at this point its energy has again become entirely potential.

148. The Period of the Pendulum. We have now to investigate the quantities upon which the time of vibration of the pendulum depends. Since the pendulum vibrates with harmonic motion, its time of vibration T is the same as the period of a particle executing harmonic motion, that is (p. 74),

$$T = 2\pi\sqrt{-\frac{\text{displacement}}{\text{acceleration}}}.$$

The displacement of the pendulum at any point, Fig. 97, is the length of the arc from the position of equilibrium to that point, or

$$\text{displacement} = -l\theta,$$

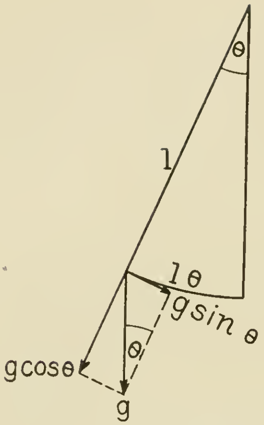

FIG. 97. The acceleration of the pendulum is a component of g.

where θ is the angle of displacement. The total acceleration g, acting on the bob, may be resolved into two components, one of which, $g \sin \theta$, acts in the direction in which the bob is moving, thus changing its motion, while the other, $g \cos \theta$, acts at right angles to this direction, and therefore does not influence the motion. Since the displacement of the pendulum is

$$-l\theta,$$

and the only acceleration affecting its motion is

$$g \sin \theta,$$

we may substitute these values in the expression for the period given above, and obtain

$$T = 2\pi \sqrt{\frac{l\theta}{g \sin \theta}}.$$

If we confine the vibrations of the pendulum to very small amplitudes, $\sin \theta$ is so nearly equal to θ that we may again write

$$\sin \theta = \theta,$$

and the time of vibration, or the period, of the pendulum, becomes

$$T = 2\pi \sqrt{\frac{l\theta}{g\theta}},$$

or

$$T = 2\pi \sqrt{\frac{l}{g}}.$$

149. The Laws of the Pendulum. Inasmuch as the above expression for the period contains the only quantities that affect the time of vibration, the following laws of the pendulum, which were first announced by Galileo, may all be deduced from this single equation:

1. The time of vibration is directly proportional to the square root of the length of the pendulum.

2. The time of vibration is inversely proportional to the square root of the acceleration of gravity.

3. The time of vibration is independent of the mass of the pendulum, since the mass does not enter into the above expression for the period.

4. The time of vibration is independent of the amplitude when this is small, since the amplitude likewise does not enter into the above expression. This, however, is true only when the vibrations are so small that we may assume $\sin \theta = \theta$, and this approximation may be allowed only when the arc of swing does not exceed 3° or 4°.

150. Determination of the Acceleration of Gravity by the Pendulum. Since the length of the pendulum and the acceleration of gravity are the only variables upon which the period depends, and since both the length and the period may be very exactly measured, the acceleration of gravity may be determined with remarkable accuracy by the swinging of the pendulum. By re-writing the expression for the period,

$$T = 2\pi \sqrt{\frac{l}{g}},$$

we obtain

$$g = \frac{4\pi^2 l}{T^2}.$$

Because of the extreme accuracy of this method, the pendulum is carried from place to place on the surface of the earth, and the value of the acceleration of gravity is determined at many points. When these measurements are arranged with respect to latitude, allowances being made for the variations due to elevation above the sea-level and for the centrifugal force, the increase of gravity towards the poles is at once observed, and a correct estimate of the form of the earth is obtained.

151. Sympathetic Vibrations. A most interesting phenomenon connected with the swinging of the pendulum is the fact that one pendulum, as it vibrates, tends to set in vibration another pendulum of the same period, if the two are mounted upon the same support. This phenomenon, which is known as that of sympathetic vibrations, may be explained as follows: The vibrating pendulum communicates a very small movement to the support, so that this also moves back and forth through an extremely small amplitude. This movement of the support is thus transmitted to the second pendulum which hangs from it, and so this pendulum likewise receives a motion forward and back. After the second pendulum has completed its small swing and is now just about to move forward again, a second impulse comes, tending to make it move forward a second time, and thus increase the motion already imparted to it. This process will continue, a new impulse coming each time the second pendulum is about to move forward, until, in this way, the first pendulum has communicated a very considerable motion to the second. Inasmuch, however, as the law of the conservation of energy must hold true, the energy expended in setting the second pendulum in vibration must be taken from the store of energy belonging to the first. Hence, as the motion of the second pendulum increases, that of the first decreases, until the second pendulum has acquired the entire motion of the first, when the first will be brought to rest. This whole process will then be repeated in the reverse order, the second pendulum setting the first in vibration and ultimately coming to rest itself.

152. The Compound Pendulum. What has been said thus far refers more especially to the mathematical, or simple, pendulum, while the pendulums with which we have actually to deal are always compound. Nevertheless, it is most interesting to consider that the compound pendulum may be used for every purpose for which the simple pendulum could be employed; and even the expression for the time of vibration may be applied without change to the compound pendulum. This application, however, is possible only when we consider very carefully what is meant by the length of the compound pendulum, for evidently the definition of the length of the simple pendulum may in no way be extended to the compound. Thus it becomes necessary to determine in the compound pendulum a distance that is exactly equal to the length of the equivalent simple pendulum, that is, to the length of a simple pendulum that would vibrate in the same time as the compound pendulum. This distance, which is known as *the equivalent length of the compound pendulum,* may be determined by the following consideration:

153. The Center of Oscillation of the Compound Pendulum. As the compound pendulum vibrates, each of its particles may be considered as a simple pendulum tending to vibrate independently, those near the point of support striving to vibrate more rapidly than the pendulum, and those farther away having a natural period which is greater than that of the pendulum. It is obvious, therefore, that there must be within the pendulum one particle that tends to vibrate neither faster nor slower than the pendulum vibrates as a whole, that is, one having the same period as the pendulum itself. In other words, if this were the only particle in the pendulum, or if the entire mass of the latter were concentrated at this point, and this point were then connected with the support by a thread without weight, its period of vibration would be the same as that of the actual pendulum. This would then be a simple pendulum, the length of which would be the distance from the point of support to this point, while the period would be the same as that of this compound pendulum.

A point such as this exists in every compound pendulum, and is called the *center of oscillation.* Accordingly, the center of oscillation of a compound pendulum is that point within the pendulum at which the entire mass of the pendulum might be concentrated without changing the time of vibration. From this we see that the length of the compound pendulum is the length of the equivalent simple pendulum, and this is the distance from the point of support to the center of oscillation. It is this distance, then, that is used as the length of the pendulum in the expression for the time of vibration, the period of this equivalent simple pendulum being the same as the observed period of the compound pendulum.

154. The Reversibility of the Compound Pendulum. It remains, therefore, simply to determine the distance from the point of suspension to the center of oscillation, or, in other words, to locate the exact point at which the center of oscillation is situated. In the solution of this problem, we are aided by a remarkable property of the compound pendulum, namely, the interchangeability of the point of support and the center of oscillation. That is, if the pendulum is reversed, and suspended from an axis placed at its center of oscillation, so that the former center of oscillation becomes the new point of support, the former point of support will then be found at the new center of oscillation, and the pendulum so reversed will vibrate in the same time as it did in its original position. In a compound pendulum therefore, *the point of support and the center of oscillation are interchangeable;* and when the pendulum is supported from one of these points the center of oscillation will be found at the other.

The practical method, therefore, of determining the exact position of the center of oscillation is to find a point in the compound pendulum, dynamically different from the point of support, such that when the pendulum is reversed and is hung from this point, Fig. 98, it will vibrate with unchanged period. This point, when so found, is the center of oscillation of the compound pendulum; and the distance between this point and the point of support is the length of the equivalent simple pendulum, that is, it is the length of a simple pendulum that would have the same period of vibration as the

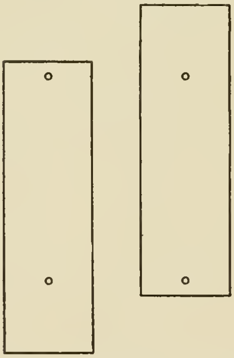

FIG. 98. The point of support and the center of oscillation are interchangeable.

compound pendulum. Thus the compound pendulum may be used instead of the simple pendulum, the length of the compound pendulum being the distance between the point of support and the center of oscillation.

155. Rotation of the Earth Shown by the Pendulum—The Foucault Pendulum. Since the pendulum is a body rotating about its point of support, the moment of inertia of the pendulum tends to maintain unchanged the conditions of rotation, and one of these conditions is the plane in which the pendulum vibrates. A pendulum, therefore, when once started, tends to continue to vibrate in the same plane, that is, it tends to maintain its original plane of vibration. Accordingly, if a pendulum were hung above the north pole of the earth and set in vibration, it would remain in the plane of vibration in which it was started, and the earth would turn under it. Since the earth revolves from west to east, the plane in which the pendulum vibrates would appear to rotate from east to west, making one complete revolution in twenty-four hours. At the equator, the pendulum would simply be carried around with the earth, and no such apparent rotation of its plane of vibration would exist, since the earth has no component of rotation about a vertical axis at the equator.

At a point intermediate between the pole and the equator, we must regard the rotation of the earth as a vector quantity, and must consider this rotation as resolved into two components, one about a vertical axis through the point chosen, and the other about an axis at right angles to the first. Thus, in Fig. 99, if *NS* represents the axis of the earth, and *M* the point on the surface at which the pendulum is suspended, the line *ON*, which coincides in direction with the axis, may be taken to represent the diurnal revolution of the earth. This actual rotation may

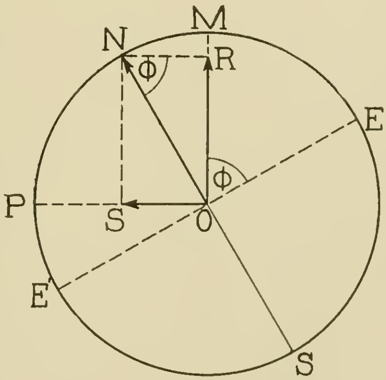

Fig. 99. Components of the earth's rotation.

be resolved into two component rotations, one OR about a vertical axis at M, and the other OS about an axis at P, the former of these, or OR, being the only component of the earth's rotation that affects the pendulum at M. If ϕ represents the latitude of M, it is seen that

$$OR = ON \sin \phi.$$

Accordingly, if the pendulum is suspended at a point intermediate between the north pole and the equator, the plane in which it vibrates will apparently rotate from east to west, but only with the component of the earth's rotation at that point. The time, therefore, in which the plane of the pendulum will make one complete revolution will be greater than twenty-four hours, depending upon the latitude of the place. The nearer the equator the point is taken, the smaller is the component of the earth's rotation about a vertical axis through that point, and the longer is the time required for the plane of the pendulum to make one complete revolution.

Foucault of Paris was the first to perform this beautiful experiment, demonstrating the rotation of the earth, and this pendulum since that time has been known as the Foucault pendulum. Foucault also showed that, if ϕ represents the latitude of the place where the pendulum is hung, the time t required for the plane of the pendulum to make one complete revolution is

$$t = \frac{24 \text{ hours}}{\sin \phi}.$$

FLUIDS

CHAPTER XI.

LIQUIDS.

156. The Characteristics of a Fluid. In the previous consideration of solid bodies, we have assumed that the particles of which these bodies are composed have a fixed, or definite position with relation to one another, and that under ordinary conditions is impossible for them to change their positions with respect to other adjacent particles. There is, however, a large class of bodies, called fluids, the particles of which are not rigidly attached to one another, but are free to move about among themselves whenever a force acts upon them. The smaller the resistance these particles encounter as they move, or attempt to move, the more nearly perfect is the fluid. A perfect fluid, therefore, is one the particles of which are free to move relatively to one another in obedience to the slightest unbalanced force.

Practically, however, it is found that all known fluids offer a certain resistance, not unlike friction, to the motion of their particles, and hence no existing fluid may be regarded as perfect. In some fluids, the cohesive force between the particles exists to a less degree than in others, and so bodies differ in their degree of fluidity. Water is a fairly fluid substance, alcohol is more completely so, and ether conforms still more nearly to the definition of the perfect fluid. Molasses, sticky oils, shoemaker's wax, etc., are examples of the more viscous fluids. In stamping a coin, the metal becomes partially fluid under the excessive pressure, and its particles therefore move with reference to one another; that is, for the instant of time during which the pressure acts, the metal becomes semi-fluid, and actually flows into the deeper portions of the die. So, also, the lower rocks in the crust of the earth, under the pressure due to the weight of the overlying strata, are folded and deformed, and become exceedingly viscous fluids, slowly yielding and flowing in obedience to the enormous pressure upon them.

If we apply to different bodies the definition of a fluid here given, we shall find that all fluids may be classified broadly into two great divisions, namely, liquids and gases. Each of these groups will be considered in turn, and the characteristics and laws they have in common, as well as the respects in which they differ, will be pointed out.

157. The Interior Pressure is Equal in all Directions. If we consider a small portion of liquid in the midst of a larger body of liquid at rest, Fig. 100, we shall find that this little portion is in equilibrium, even when a pressure is applied to the entire free surface. If we now imagine for a moment this little volume to become hollow, we know that the liquid will flow in from all sides to fill this volume, and thus we know that a pressure exists upon its sides, tending to force the liquid inward. This same pressure, therefore, must have existed before the little volume became hollow; but as this volume did not move, we see that the pressure to which it was subjected in one direction must have been just counterbalanced by an equal pressure in the opposite direction. Accordingly, since the particles of a liquid at rest do not move with respect to one another, it must follow that the pressure upon each little particle must be the same in all directions.

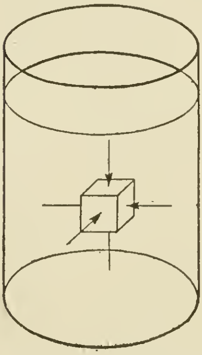

FIG. 100. The pressure in a liquid is equal in all directions.

By the term *pressure* applied to a liquid or a gas is meant *the intensity of the force acting on unit area.*

158. The Pressure is Normal to the Sides of the Containing Vessel. The great mobility of the particles of a liquid leads us, furthermore, to the important conclusion that the pressure exerted by a liquid upon the walls of the containing vessel must always be perpendicular to the sides of the vessel. If this pressure were not normal to the sides, it would be in some other direction, as *ab* Fig. 101. This oblique pressure could then be resolved into two components, one *ac,* perpendicular to the sides of the vessel, and the other *ad,* parallel to these sides. The component *ac* could produce no motion, since it is counterbalanced by an equal pressure in the opposite direction, due to the rigidity of the vessel; but the unbalanced component *ad* would produce a motion of the particle of the liquid in the direction from *a* to *d*. Since the particles of a liquid do not move along the walls of the containing vessel, it is obvious that there can be no component of the pressure parallel to these walls, and hence the entire pressure must be perpendicular. Therefore, the pressure exerted by a liquid upon the sides of the containing vessel is everywhere perpendicular to these sides.

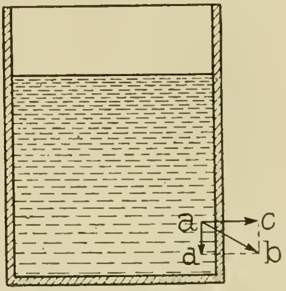

Fig. 101. The pressure is perpendicular to the sides of the vessel.

159. External Pressure—Pascal's Law. When the result of any pressure upon a liquid is considered, it is found that there are two conditions under which this pressure may exist; namely, the pressure applied to a liquid from some external source, and the pressure upon the particles of a liquid due to the weight of the liquid itself. Each of these conditions will be considered in turn.

In investigating the application of an external pressure upon a liquid, it is usual to neglect the weight of the liquid, that is, we may consider the liquid with which we are dealing as weightless, since, in general, its weight is small as compared with the intensity of the pressure.

If we imagine a closed reservoir filled with liquid, and a little cylinder provided with a piston inserted in this reservoir, a pressure applied to the piston will be transmitted undiminished to the particles of liquid immediately beneath. As these particles attempt to move, they exert a pressure upon the particles, adjacent to them, these press upon other particles, and so on, until this pressure is transmitted undiminished to every portion of the reservoir equal in area to the area of the piston. This principle of the transmission of pressure, which is true for gases as well as for liquids, was first announced by Pascal, and is known as Pascal's law. *Pascal's law declares, therefore, that the pressure applied to a fluid is transmitted equally in all directions.*

If the reservoir is provided with another cylinder and another piston similar to the first and of equal area, Fig. 102, this second piston will be forced out with a pressure equal to the pressure applied to the first. If a third cylinder and piston similar in every way to the second is placed adjacent to the latter, a pressure will also be developed upon this piston equal to the pressure applied to the first. If these two pistons are now united, the result will be a single piston of double the area of the first, and upon this larger piston there will be developed a total pressure twice as great as that exerted upon the first. By this means a small pressure exerted upon a small piston may be made to develop a large pressure upon a large piston, the ratio of these two pressures being the ratio of the areas of the two pistons, as in the hydrostatic press.

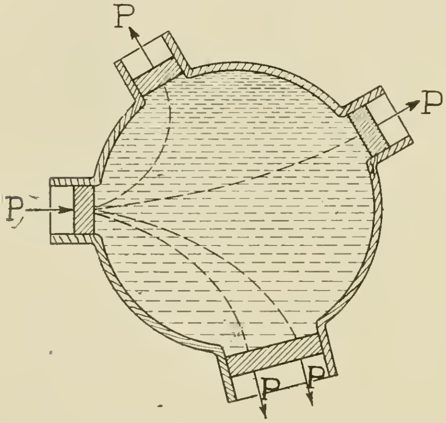

FIG. 102. The pressure is transmitted equally in all directions—Pascal's law.

160. The Principle of the Hydrostatic Press. The hydrostatic press consists of a large cylinder connected with a small one, each filled with a liquid, and each provided with a piston, Fig. 103. If a pressure is applied to the small piston, this pressure will be transmitted undiminished to every equal area through the liquid. The total pressure upon the larger piston will then be as many times greater than that applied to the smaller as the area of the larger piston exceeds that of the smaller. This is an illustration of the principle already announced (p. 123), that a small force is able to create a very large force, and that there is no such law as the conservation of force.

The *work* done by the small force acting upon the small piston, however, is in every respect equal to that performed by the larger force upon the larger piston, and the law of the conservation of energy is obeyed. In other words, the product of the force applied to the smaller piston and the distance through which this force moves is equal to the product of the force acting upon the larger piston and the distance through which it moves. It follows, therefore, that the distances through which the pistons move are inversely as their areas.

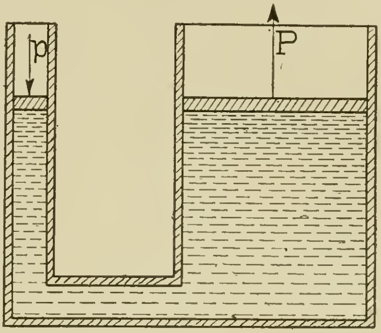

FIG. 103. The principle of the hydrostatic press.

161. Pressure due to Gravity—The Surface of a Liquid is Normal to the Resultant Force. When we now consider the pressure upon a liquid due to gravity, that is, to its own weight, the first point apparent is the fact that the surface of the liquid is horizontal. If this were not so, it would have in some other form, as for instance, that shown in Fig. 104. A particle of liquid at the surface is acted upon by the force of gravity, and this force may be resolved into two components, one acting parallel to the surface, and the other perpendicular thereto, and of these, the component parallel to the surface would cause the particle of liquid to move. Inasmuch, however, as the surface of a liquid does not move, we see that there can be no component of the force of gravity acting in this direction, and hence the entire force of

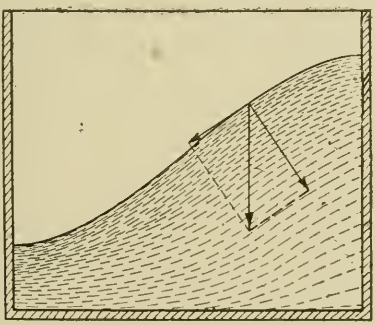

FIG. 104. The free surface of a liquid at rest is horizontal.

gravity must act perpendicularly to the surface, or, in other words, the free surface of a liquid at rest is horizontal.

A beautiful illustration of the principle that the surface of a liquid in equilibrium is always perpendicular to the resultant of all the forces acting upon it, may be seen in the form assumed by the surface of water in a rotating vessel, Fig. 105. Each particle of the liquid is then acted upon by two forces: one, the force of gravity acting vertically downward; and the other, the centrifugal force acting horizontally outward, the latter being equal and opposite to the centripetal force that causes the liquid to move in its circular path. The vertical force due to gravity is the same upon all equal particles, but the horizontal centrifugal force is different at different points, becoming greater as the distance of the particles from the axis of rotation increases. These two forces combine to form a resultant, and this resultant, which is different at different points both in direction and magnitude, is everywhere perpendicular to the surface. The surface then assumes a parabolic form, and is thus in complete equilibrium, being everywhere perpendicular to the resultant of the forces acting.

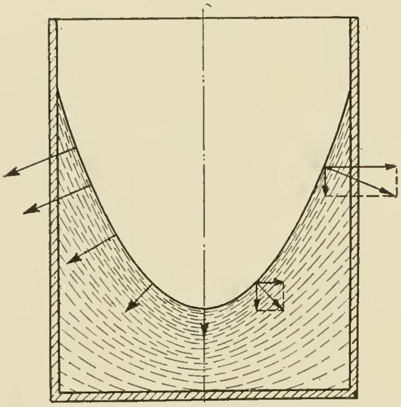

Fig. 105. The surface of a liquid is perpendicular to the resultant force.

162. The Pressure is Proportional to the Depth. When a liquid is at rest, a particle beneath the surface must sustain the weight of the column of liquid above it; and so, if two particles are taken at different depths, the pressure upon the lower one is greater than that upon the upper. From this principle is derived the law that *the pressure within a liquid is proportional to the depth.*

If two particles are taken at the same depth, Fig. 106, the height of the column of liquid above each particle is the same and hence the pressure at one of these points must be the same as that at the other. From this consideration is derived the further law that *the pressure at all points of the same horizontal level is the same.* If the pressure were not the same at points of the same horizontal level, a flow of the liquid would take place until the height of the column of liquid above one point were the same as that above another; that is, until the surface of the liquid were horizontal. A fountain, therefore, will rise, friction and the resistance of the air being neglected, until the water reaches the level of the original source. Upon this principle depends also the action of the water level.

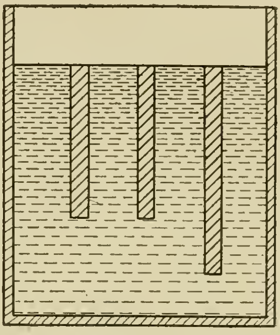

Fig. 106 The pressure increases with the depth.

163. Pressure upon the Base of a Vessel Filled with Liquid. It is now necessary to determine the pressure due to gravity upon the base of a vessel filled with liquid. In the solution of this problem, it is important to consider three cases, depending upon the inclination of the walls of the vessel to the base; 1, when the walls of the vessel are vertical; 2, when the walls are flaring outward, the vessel being wider at the top than at the base; 3, when the walls are contracted towards the top, the vessel then being narrower at the top than at the base.

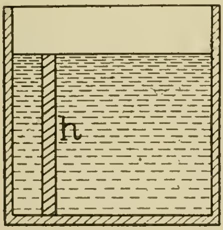

Fig. 107. With vertical sides the pressure on the base is the actual weight of the liquid.

Let us assume in the first instance, that the walls of the vessel are vertical, Fig. 107, that the base has an area of A square centimeters, that the depth of the liquid is h centimeters, and that the weight of one cubic centimeter of the liquid is w grams. Each square centimeter of the base must then sustain the actual weight of the column of liquid above it, or hw grams. Therefore the entire base, which has an area of A square centimeters, must sustain a total pressure A times as great, or Ahw grams. But this is the actual weight of the liquid contained in the vessel. Accordingly, in a vessel with vertical sides, the pressure upon the base, due to the weight of the liquid is equal to the weight of the liquid contained.

In the second instance, we have to deal with a vessel, the sides of which diverge towards the top, so that the vessel is wider at the top than at the base. Here again, if the area of the base and the depth of the liquid are the same as before, the pressure upon one square centimeter of the base is hw grams, and therefore, the total pressure upon the entire base is Ahw grams, or exactly what the pressure upon the base would have been if the vessel had had vertical sides.

The truth of this will appear the more obvious if we consider that the pressure P' of the liquid at any point against the flaring sides of the vessel, Fig. 108, is equal to the weight of the column of liquid one square centimeter in cross-section vertically above that point, and that this pressure is everywhere perpendicular to the sides. Because of the rigidity of the vessel, the sides exert an equal pressure P in the opposite direction upon the adjacent particles of the liquid. The latter pressure may be resolved at each point into a horizontal and a vertical component, each horizontal component H holding in equilibrium an equal horizontal component from the opposite side of the vessel, and each vertical component V sustaining the vertical component of the pressure due to the weight of the liquid directly above it.

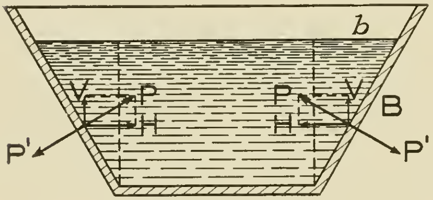

FIG. 108. The flaring sides sustain the weight of the liquid above them.

If these vertical components are summed up for the entire area of the sloping sides, it will be found that their sum is exactly equal to the weight of the liquid immediately above these sides, as may be seen from the following demonstration:

Remembering that the pressure P is the *force per unit area* exerted by the sides, and that V is therefore the vertical component of this force *per unit area,* we see that the total upward pressure which the sides exert is

$$B\,V,$$

where B denotes the area of the sloping sides. Moreover, if b is the area of the horizontal liquid surface vertically above these sides, we have, from similar triangles,

$$\frac{V}{P}=\frac{b}{B},$$

or

$$B\,V=b\,P.$$

But $B\,V$ is the total vertical pressure which the sides exert, and $b\,P$ is the weight of the liquid above these sides, since b is the number of square centimeters in the surface of this liquid. and P is the weight of the column below each square centimeter. It follows, therefore that the sloping sides exert an upward pressure upon the liquid equal to the weight of the liquid immediately above them. In other words, the sloping sides uphold the weight of the liquid vertically above these sides, so that the only pressure which the base sustains is the weight of the liquid vertically above the base. The pressure upon the base, therefore, is exactly the same as it would be if the vessel had vertical sides; that is, it is the weight of a column of the liquid, the cross-section of which is the area of the base, and the height of which is the depth of the liquid. Moreover, the total pressure upon the base of a vessel with flaring sides, due to the weight of the liquid, is thus seen to be less than the weight of the liquid contained in the vessel.

If, in the third instance, the sides of the vessel contract towards the top, we shall see that the pressure upon the base, A, is again the same as it would have been if the sides had been vertical, and the vessel had been filled with liquid to the same depth h. This will be at once understood, Fig. 109, if we select a square centimeter a, under the sloping sides of the vessel, and investigate the pressure that actually exists upon it. The total downward pressure upon the area a is the weight of the column of liquid *ab* vertically above it, plus whatever downward pressure exists at *b*. At *b* the pressure must be the same as the pressure at *c*, since *b* and *c* are points at the same horizontal level. The pressure at *c*, and therefore also the pressure at *b*, is the weight of the column of liquid *cd*. Accordingly, the pressure at *a* is the weight of the column of liquid *ab* plus the weight of the column of liquid *cd*, that is, $(ab + cd)w$, or hw. A similar analysis may be applied to each square centimeter of the base, from which it is evident that the entire pressure upon the base of A square centimeters is Ahw, or exactly what it would have been if the sides had been vertical.

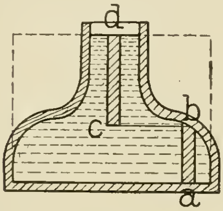

FIG. 109. Pressure on the base greater than the weight of the liquid.

We may perhaps see this more exactly by reference to Fig. 110, in which, as before, P' represents the pressure of the liquid at a given point against the sloping sides, and P is the equal and opposite pressure which the sides exert upon the liquid P is therefore equivalent in magnitude to the weight of a column of the liquid 1 square centimeter in cross-section, and of a height equal to the depth of the liquid at that point. As before, P is resolved into a horizontal component H and a vertical component V, the horizontal component being held in equilibrium by an equal component from the opposite side, and the vertical component V being transmitted to the bottom of the vessel.

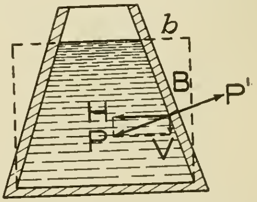

FIG. 110. The sloping sides transmit a pressure to the base.

Since this is the pressure that is transmitted from one square centimeter of the side, it is evident that

$$B V$$

represents the total pressure which the entire sides transmit to the base, where B is the number of square centimeters in the area of the sides. If b is the horizontal area vertically over the sides at the level of the liquid, we have, by similar triangles,

$$\frac{V}{P} = \frac{b}{B},$$

or

$$B\,V = b\,P.$$

Here $B\,V$, as shown above, is the pressure which the sides transmit to the base, and $b\,P$ is the weight of the liquid which would fill the volume above the sloping sides if the walls of the vessel had been vertical. This is true since b is the number of square centimeters of additional surface which the liquid would then have had, and P is the weight of the column of the liquid below each of these square centimeters.

In other words, the sloping sides of the vessel exert upon the liquid a downward pressure which is transmitted to the base of the vessel, this pressure being equal to the weight of the liquid which would have been above the present sides, if the walls of the vessel had been vertical and not sloping. From this it follows that, in the case of a vessel contracted towards the top, the pressure upon the base, due to the weight of the liquid is equal to the pressure which would have existed if the vessel had had vertical sides, and is therefore greater than the weight of the liquid which the vessel actually contains.

If these three vessels have bases of equal area, and are filled with the same liquid to the same depth, the pressure upon the base of one is exactly equal to the pressure upon the base of either of the others. In all cases, therefore, the pressure upon the base of a vessel filled with liquid is equal to the weight of a column of the liquid, the cross-section of which is equal to the area of the base and the height of which is equal to the depth of the liquid.

The pressure, therefore, upon the base of a vessel, due to the weight of the liquid contained, may be equal to the weight of the liquid, may be less than the weight of the liquid, or may be greater than the weight of the liquid, depending upon the shape of the vessel. This is known as the hydrostatic paradox.

164. The Upward Pressure in a Fluid. In the previous pages, it has been seen that when liquids are influenced by gravity alone, every liquid particle in the midst of a body of liquid must sustain a pressure equal to the weight of the column of liquid above it. The pressure upon any liquid area beneath the surface of a fluid is then equal to the weight of a column of liquid, the cross-section of which is the area considered, and the height of which is the depth of this area beneath the surface. But as this little imaginary liquid area is in equilibrium, it is evident that this downward pressure must be counterbalanced by an equal upward pressure. In fact, since the particles of the liquid are everywhere in equilibrium, the pressure in one direction must be equal to the pressure in every other direction, and hence there must be also a pressure upward. This pressure upward becomes apparent when the downward pressure is removed, Fig. 111, as when a cardboard disc, held over the open end of a tube, is immersed in the liquid. The pressure of the liquid upward against the cardboard disc will press the disc against the open end of the tube, and prevent the liquid from entering.

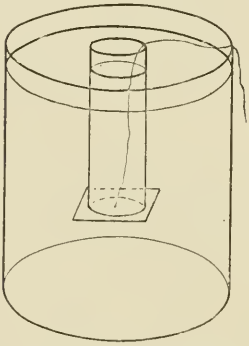
FIG. 111. The upward pressure in a liquid.

165. The Principle of Archimedes. If a body in the form of a parallelopiped is entirely immersed in a liquid, the upper surface of the body being parallel to the surface of the liquid, the six faces of the body will experience pressures from the liquid. The pressures upon the four lateral faces, however, need not be considered, since the pressure upon one face in one direction is exactly counterbalanced by an equal pressure upon the opposite face in the opposite direction, these four lateral faces all being at the same depth. It remains, therefore, simply to consider the pressures upon the upper and the lower faces of the parallelopiped.

The pressure acting downwards upon the upper face of this body is the weight of a column of liquid in the cross-section of which is equal to the area of this face, and the height of which is equal to its depth beneath the surface, Fig. 112. If the edges of the parallelopiped are a and b centimeters in length, and c centimeters in height, the area of the upper face and also of the lower face is ab square centimeters. If the depth of the liquid above the upper surface is h centimeters, and the weight of one cubic centimeter of the liquid is w, the pressure acting downward upon the upper face of the body is $abhw$ grams. The pressure acting upward upon the lower face is the weight of a column of liquid, the cross-section of which is the area of this face, and the height of which is its depth beneath the surface. As this face is at a depth of $h + c$ centimeters the upward pressure upon this lower face is $ab(h + c)w$ grams. The resulting upward pressure upon the parallelopiped is the difference between this upward pressure upon the lower face and the downward pressure upon the upper face, that is

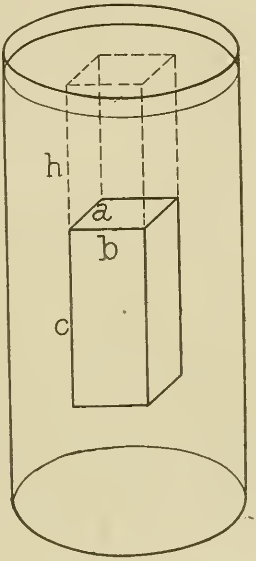

FIG. 112. Upward pressure equal to weight of displaced liquid.

$$ab(h + c)w - abhw = abcw \text{ grams.}$$

But abc is the volume of the immersed parallelopiped, and therefore is also the volume of the liquid displaced by this body. Moreover, as w is the weight of one cubic centimeter of the liquid, the expression

$$abcw \text{ grams}$$

is the weight of the liquid displaced by the parallelopiped. Hence the resultant upward pressure upon the parallelopiped immersed in the liquid is equal to the weight of the liquid displaced.

If an irregular body is immersed in the liquid, it may be regarded as divided into a series of very small parallelopipeds, for each of which the above reasoning holds true. Hence it is, that any irregular solid immersed in a liquid experiences a resultant upward pressure equal to the weight of the liquid displaced. This effect is known as the Principle of Archimedes, and may be stated in these words: *A body immersed in a liquid is buoyed upward with a force equal to the weight of the liquid displaced.*

166. The Cylinder and Bucket Experiment. The Principle of Archimedes, here described, may be illustrated by the "cylinder and bucket" experiment. A metal cylinder and a cylindrical bucket that will exactly contain this cylinder are suspended from the beam of a balance, and counterpoised by weights in the other pan. If the cylinder is then hung beneath the bucket and immersed in a liquid, Fig. 113, the balance is no longer in equilibrium, inasmuch as the cylinder experiences an upward pressure equal to the weight of the liquid displaced. Since it displaces an amount of liquid equal to its own volume, and therefore to the volume of the bucket, if the bucket is now exactly filled with the liquid, the equilibrium of the balance is again restored.

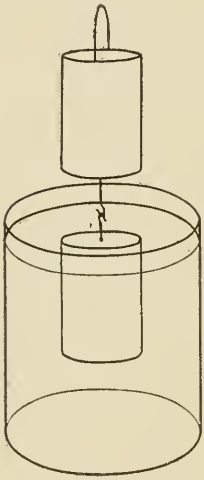

FIG. 113. The cylinder and bucket experiment.

167. Loss of Weight Transmitted to Bottom of Containing Vessel. From the above experiment it would seem that when a body is immersed in a liquid it apparently loses in weight an amount equal to the weight of the liquid displaced. This loss of weight, however, is only apparent, as may be seen by considering the vessel in which the liquid is contained. As soon as the body is immersed, the surface of the liquid rises, and the pressure upon the bottom of the vessel is increased, just as if a quantity of the liquid equal to the volume of the immersed body had been added to the vessel. The weight which the immersed body apparently loses, therefore, is transmitted to the bottom of the vessel containing the liquid. This might have been anticipated from the fact that action and reaction are equal and in the opposite direction; for the upward action of the liquid upon the immersed body must be equal to the downward reaction of the immersed body upon the liquid.

168. Floating Bodies. If W equals the weight of the body immersed, P equals the upward pressure of the liquid upon the body, and F equals the resultant of these two forces acting downward, then

$$F = W - P.$$

If the weight of the body is greater than the upward pressure, then W is larger than P, the force F is positive and the body sinks. A body therefore, will sink in a liquid when its weight is greater than the weight of the liquid it displaces. If the weight of the body is equal to the weight of the liquid displaced, then W is equal to P, the resultant force F is then zero, and the body will remain in equilibrium wherever it may be placed beneath the surface. If the weight of the body is less than the weight of the liquid displaced, then P is larger than W, the force F is negative, and the body will float upon the surface. A body, therefore, will float upon a liquid whenever the weight of the body is less than the weight of the liquid displaced. Furthermore, a body placed upon the surface of a liquid will sink until the weight of the liquid displaced is equal to the weight of the body. Thus it is that iron will float upon mercury as cork floats upon water; the iron sinks in the mercury until it displaces an amount of the mercury equal in weight to the weight of the iron. A balloon will rise in the air because it displaces an amount of air greater in weight than the weight of the balloon.

169. Specific Gravity. Density has been defined (p. 77) as the ratio of mass to volume that is, it is the mass contained in unit volume. The principle of Archimedes gives us a means of determining density, for here we have a method of ascertaining the volume of a body, however, irregular in shape the body may be. If a body is weighed in air and then weighed immersed in a liquid, the apparent loss of weight in the latter case is equal to the weight of the liquid displaced. If this liquid is water, the number of grams the body apparently loses in weight is numerically equal to the number of cubic centimeters in the volume of the body. The mass and the volume of the body being now known, the density is expressed simply as the ratio of the former to the latter.

The practical method of determining density is then as follows. The substance is first weighed in air, then weighed immersed in water. The difference between these two weights is the number of grams of water displaced, and this is numerically equal to the number of cubic centimeters in the volume of the body. The ratio of the weight of a body in air to its loss of weight in water is, therefore, numerically the same as the ratio of the mass of the body to its volume, that is, numerically the same as the density. This agreement, however, is true only because the density of water is unity, that is, because each cubic centimeter of water contains the mass of one gram. The number of grams a body apparently loses when immersed in water therefore, is, the same as the mass of the water displaced. Accordingly, the ratio of the weight of the body in air to its loss of weight in water is really the ratio of the mass of the body to the mass of an equal volume of water, that is, it is the ratio of the density of the body to the density of water.

Density determined in this way, therefore, is not absolute, but relative to the density of water, and is known as specific density, or, more generally, as specific gravity. Density and specific gravity are therefore different in meaning, since density is simply the ratio of mass to volume, while specific gravity is the ratio of the density of a substance to the density of water. It is, accordingly, only because the density of water is unity that density and specific gravity are numerically equal.

170. The Hydrometer. The instrument known as the hydrometer furnishes another means of measuring relative densities, or specific gravities, although in action it is fundamentally different from the above method by the balance. The method as described above, namely, that of weighing the body in air and then in water, consists in determining *the masses of equal volumes;* the action of the hydrometer, however, depends upon the principle of determining *the volumes of equal masses.*

The hydrometer consists of a glass cylinder, Fig. 114, weighted at its lower end, and surmounted by a graduated tube. Mercury is then added, until the cylinder, as it floats in pure water, sinks to the point marked zero, when it displaces its own weight of water. If it is then placed in a liquid lighter than water, it will sink deeper, and if placed in a liquid heavier than water, it will not sink so deep, before it displaces a volume of the liquid equal to its own weight. The weights of the displaced liquids, therefore, remain always the same, and equal to the weight of the instrument itself; while the variations in the volumes of these displaced liquids are indicated by the different depths to which the instrument sinks. By a proper calibration of the tube, the densities of the liquids, therefore, may be read at once.

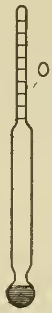

FIG. 114. The hydrometer.

171. The Flow of Liquids Through an Orifice—Torricelli's Theorem. If an opening is made in the bottom of a vessel filled with liquid, Fig. 115, the liquid will flow from the orifice, because of the unbalanced pressure due to the weight of the liquid within. It follows, therefore, (p. 121), that the kinetic energy of each particle of the moving liquid jet is

$$\frac{m v^2}{2},$$

where m is the mass of the particle, and v is the velocity with which it leaves the opening. In filling the vessel originally with the liquid, the weight of this particle, or mg, had to be lifted through the height h, equal to the depth of the liquid, and an

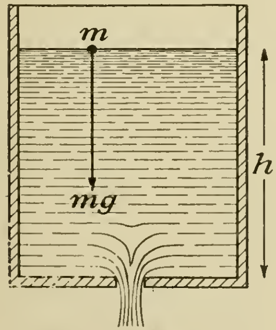

FIG. 155. The flow of liquid—Torricelli's theorem.

amount of work equal to mgh was done upon the particle in giving it its position at the surface. Accordingly, (p. 120), the potential energy of a particle of the liquid at the surface, before the flow takes place, is

$$mgh.$$

While the particle that flows out is situated at the depth of the orifice and not at the surface, the condition of the liquid remaining in the vessel is the same as if this particle had been removed from the surface. The liquid remaining in the vessel, therefore, has lost in potential energy an amount equal to

$$mgh,$$

and this loss of potential energy is equal to the kinetic energy acquired by the outflowing particle, and therefore

$$mgh = \frac{m v^2}{2}.$$

From this

$$v^2 = 2gh,$$

or

$$v = \sqrt{2gh}.$$

From this equation it is seen that the velocity with which a liquid issues from an orifice depends only upon the depth of the liquid and upon the acceleration of gravity.

The remarkable truth apparent from this expression is the fact that the velocity of the outflowing liquid is entirely independent of the mass of the liquid, so that a vessel filled with mercury will empty itself no more quickly than will a vessel filled with water, provided both are filled to the same depth and have equal orifices at the bottom. In explaining this unexpected result, we must remember that although the force acting upon the mercury is thirteen and a half times greater than the force acting upon the water, the mass, and therefore, the inertia, of the mercury are also thirteen and a half times greater than the mass and the inertia of the water, hence (p. 80), the acceleration given to one just equals that imparted to the other, and these two liquids will flow from their respective vessels in equal times.

This equation also shows that the velocity of the outflowing liquid is proportional to the square root of the depth, Fig. 116, that is, if the velocity of a liquid flowing from an orifice at a certain depth is a certain amount, the velocity from an orifice at a point four times as deep will be only twice that amount. This expression for the velocity with which a liquid flows from an orifice is also seen to be the same as that for the velocity acquired by a body in falling freely through a vertical height h (equation 16, p. 35). A liquid, therefore, will flow from an orifice in the bottom of a vessel with the same velocity that a body would acquire in falling freely through a vertical height equal to the depth of the liquid. This is known as Torricelli's Theorem.

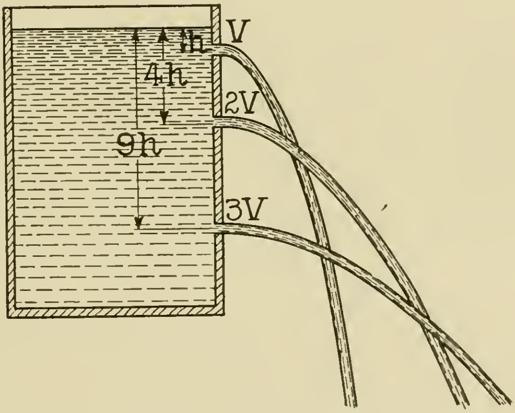

Fig. 116. The velocity of flow is proportional to the square root of the depth.

172. Lines of Flow—The Contracted Vein. Another phenomenon connected with an outflowing stream of liquid may be readily observed, especially if the wall of the vessel containing the orifice is thin, that is, if the flow takes place from an opening in a thin wall. While the particles of the liquid directly above the orifice move vertically downward towards this opening, Fig. 117, the particles near the sides and bottom of the vessel have a lateral motion as they approach the orifice. The paths of all these particles are called the lines of flow, and these are different in direction in different parts of the vessel.

The lateral motion thus acquired by the particles, as they move along the lines of flow in the vessel, cannot instantly be overcome when the orifice is reached, and so the particles pass through the opening forming lines of flow in the jet that are not parallel, but slightly convergent. On this account, the issuing column of liquid suffers a certain contraction a little distance below the opening, known as the contracted vein, or the *vena contracta*. The volume of liquid, therefore, that flows in a given time from an opening in the thin wall of a vessel is proportional, not to the size of the opening, but to the cross-section of this contracted vein at its narrowest point, and this is often not more than six-tenths of the size of the opening. This lateral movement of the liquid particles may be partially prevented by the addition of a short tube, or ajutage, of the same diameter as the opening, with the result that the flow of liquid is increased to about eight-tenths of the theoretical amount.

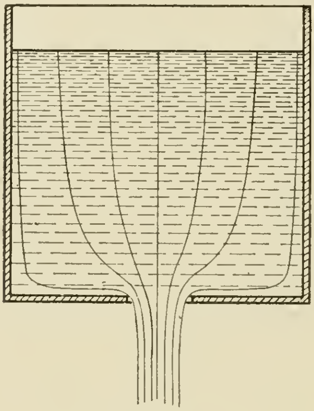

FIG. 117. Lines of flow—The contracted vein.

173. The Oscillations of a Falling Drop. For a short distance below the contracted vein the stream is continuous and practically cylindrical, after which it becomes gradually smaller in diameter as the velocities of the falling particles become greater. After a certain limiting diameter has been reached, constrictions, or waists, are formed, due to certain phenomena of the surface tension which will be studied later, and the jet then soon breaks into drops, Fig. 118. Before the drop is actually formed, however, these waist-like constrictions become more pronounced until the mass of water that will later form the drop is connected with the body of the stream only by a long, narrow neck of the liquid, and this ultimately breaks. The drop is then free, but elongated vertically; and as its particles endeavor to assume the spherical form,—this effect also being due to the action of surface tension,—they are carried by their inertia past this form, and become elongated horizontally. The same process is then repeated in the reverse direction, the drop becoming elongated first in one way and then in the other as it falls.

These oscillating drops may be conveniently studied by illuminating them with an intermittent light, that is, by placing before the source of illumination a rotating disc from which a sector has been cut. If a flash of light comes through the open sector at a time when a drop has reached a certain point, and a second flash can be made to come when the next drop has reached exactly the same point, the drop will have the appearance of standing perfectly still, and its form may be minutely examined. Another drop, a little below, will have assumed another form; and if it is possible thus to observe at the same time several drops in different phases of their vibrations, they may be seen to pass through all the variations of form that occur in a falling drop of liquid. Between each of these oscillating drops will be noticed the tiny spherical drop, into which has been contracted the elongated neck that connected the drop with the body of the jet just before the final separation of the two took place.

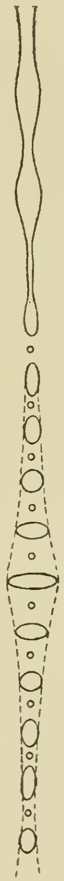

FIG. 118. Falling drops.

174. The Reaction of an Outflowing Jet. We have seen, (p. 173), that the pressure due to the weight of a liquid is equal at all points of the same horizontal level, and is everywhere perpendicular to the sides of the containing vessel. Accordingly, the pressure at any point as *A*, Fig. 119, is exactly equal and opposite to the pressure upon the opposite point, as *B*, and the vessel is in equilibrium. If an opening is made in one of the sides, as at *A*, the pressure here is then no longer directed against the wall of the vessel, but against the liquid, so that the latter is set in motion and issues as a jet. The pressure at *B*, however, remains unaffected by this change; and since this pressure is no longer held in equilibrium by an equal pressure upon the opposite wall at *A*, it becomes effective in producing motion, and the vessel, if free to move, is forced backward. This is known as the reaction of an outflowing jet, and forms an excellent illustration of Newton's third law of motion (p. 107).

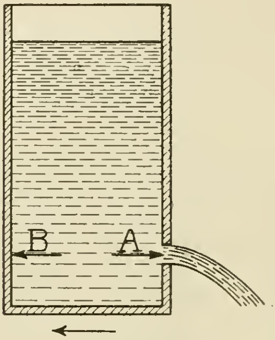

FIG. 119. The reaction of an outflowing jet.

It was, indeed, upon this principle that the first steam engine was constructed by Hero of Alexandria, in the second century, B. C. Hero generated steam in a pivoted spherical vessel, provided with tubes, or arms, bent at right angles, and from these the steam escaped. The pressure forcing out the steam produced an equal and unbalanced reaction upon the sides of the tubes opposite the jet, thus giving a backward motion to the arms, and therefore a rotary motion to the whole. The pinwheel in fireworks, the modern turbine water wheel, and Barker's water-mill, the latter most commonly seen in some varieties of lawn sprinklers, depend for their operation upon the reaction of an outflowing jet.

CHAPTER XII.

GASES.

175. Similarity between Gases and Liquids. By applying to a gas the definition of a fluid previously given (p. 167), we find that this also is a body the particles of which are free to move under the action of an unbalanced force, and hence all gases as well as liquids are fluids. While there are many ways in which these two bodies differ from each other, as will be pointed out later, there are, nevertheless, certain characteristics and laws that they have in common, due to the mobility of their particles.

Because a gas is a fluid, it must, like all other fluids, transmit pressure equally in all directions, hence Pascal's law, (p. 170) is true for gases as well as liquids. The property of elasticity of volume is a characteristic that is possessed by liquids and gases alike, for both resume their original volumes when the pressures to which they have been subjected are relieved. It is true that liquids are far less compressible than gases; but when the pressure is removed, both assume their original volumes at once and completely, so that both liquids and gases are perfectly elastic.

Another point of resemblance between the two is the fact that both have weight. By suspending a hollow ball from the beam of a balance and weighing it when filled with air and again when the air has been exhausted, the difference in weight gives the weight of a volume of air equal to the interior volume of the ball. In this way it is found that when weighed under the standard conditions to be described later, one liter of air has a mass of 1.2932 grams.

176. The Pressure of the Atmosphere. Since air is a gas, and therefore a fluid, the atmosphere about the earth may be compared to a great aerial ocean, at every point of which the same laws of pressure apply that apply to fluids in general. A litte area, therefore, in this ocean experiences a pressure due to the weight of the column of the atmosphere above it, and therefore at points near the earth the pressure must be greater than at regions more elevated. In other words, the law that the pressure increases with the depth applies to the atmosphere as to all other fluids; and because of the mobility of the particles, this pressure is equal in all directions.

The fact that the atmosphere exerts a great pressure due to its weight becomes at once apparent when the pressure of the air upon one side of a diaphragm is removed, as, for instance, when the hand is placed over the open top of a bell-jar, or when a piece of thin rubber is tied over the top of a receiver, and the air within is exhausted. The pressure downward upon this rubber membrane, due to the weight of the atmosphere, is then no longer counterbalanced by an equal pressure upon the opposite side, and the membrane is often broken by the pressure due to the weight of the air above it.

That the pressure of the air is equal in all directions, and therefore upward, may be seen by exhausting the air from the upper part of a vertical cylinder containing a piston, to which a heavy weight has been attached. Before the air above the piston is exhausted, the pressure upon its two sides is the same; but when this air has been removed, the pressure upon its upper surface is less than that upon its lower surface, hence there is a resultant pressure upward sufficient to lift the piston with the attached weight. A simpler method of illustrating the same principle consists in placing a sheet of writing paper over a tumbler full of water and inverting the whole. The water will be held in the glass by the upward pressure of the air upon the lower surface of the paper.

There is another consequence that follows from the fact that we are living in a sea of air, the particles ofwhich have weight. All bodies in the air are immersed in this gaseous fluid, and accordingly, by the principle of Archimedes (p. 180), they experience an upward pressure equal to the weight of the atmosphere displaced. From this it follows that the apparent weight of a body when weighed in the air is less than its true weight by the weight of its own volume of air.

177. The Barometer. The pressure of the air due to its weight was recognized by Galileo, but it remained for his pupil Torricelli to devise a means whereby this pressure might be measured. Torricelli conceived the idea of balancing the pressure due to the weight of the air by the weight of a column of mercury that should produce an equal pressure. This he accomplished by filling with mercury a long glass tube, closed at one end, and inverting this tube in a vessel, or cistern, of mercury, with the open end immersed beneath the surface. It was then found that the mercury fell within the tube until the height of the column remaining was 76 centimeters above the surface of the mercury in the vessel.

This simple device, which is still used for measuring the pressure of the air, forms the instrument known as the barometer. If the tube has vertical walls, the pressure at its open end is the weight of the column of mercury the tube contains. The pressure of the air, therefore, upon an area of the surface of the mercury in the vessel equal to the cross-section of the tube is also equal to the weight of this column of mercury. If the cross-section of the tube is one square centimeter, this mercury column, 76 centimeters in height, will contain 76 cubic centimeters of mercury; and as one cubic centimeter of mercury weighs 13.596 grams, the normal pressure of the atmosphere is $76 \times 13.596 = 1033.3$ grams per square centimeter. The region above the mercury column within the barometer tube is known as the Torricellian vacuum.

178. The Pressure at a Point Within the Barometer. Inasmuch as no matter exists within the Torricellian vacuum, the pressure upon the surface of the mercury within the tube at *A*, Fig. 120, is zero. At a point lower down in the tube, as at *B*, the pressure is the weight of the column of mercury *AB*, or it may be considered as equal to the pressure of one atmosphere, that is, the weight of the column *AD*, minus the weight of the column of liquid *BD* beneath it. The pressure at any point as *C* is that due to the weight of the

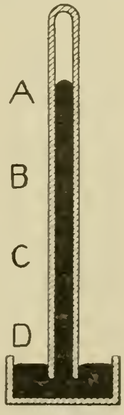

FIG. 120. The barometer.

column *AC*, or it is the pressure of one atmosphere *AD*, minus the weight of the column of mercury *CD* beneath it, etc.

179. The Work of Otto von Guericke. Contemporaneous with Torricelli in Italy, Otto von Guericke in Magdeburg, Germany, independently devised a method almost identical with Torricelli's for measuring the pressure of the air. By using a longer and larger tube, and filling it with water instead of mercury, he found that a column of water nearly 34 feet high was necessary to balance the pressure of the air. It is with Otto von Guericke, likewise, that the conception of the air-pump originated, and to him is also due the apparatus for demonstrating the pressure of the air known as the Magdeburg hemispheres. These consist of two hemispherical cups, fitting tightly together, but easily separated when filled with air. When, however, the air from within is exhausted, the pressure of the atmosphere upon their exterior surface forces these hemispheres together with a pressure of more than 1 kilogram per square centimeter of their cross-section.

180. Pressure and Density of the Air. Since the pressure in all fluids increases with the depth, the pressure of the atmosphere near the surface of the earth is greater than at an elevation above the surface, and the column of mercury in the barometer stands higher at a low altitude than at a greater one. The barometer, therefore, furnishes a very reliable means indeed for determining the height of mountains. The normal, or standard, conditions for comparing the pressures of the air are at sea-level and at a temperature 0° Centigrade; and to these standard conditions all measurements must be reduced before the varying pressures of the air may be compared.

Inasmuch as a gas is a compressible fluid, the lower strata of the atmosphere, where the pressure is greater, are denser (p. 77) than are the upper strata, where the pressure is less; that is, near the surface of the earth a larger mass of air is contained in the unit volume than is contained in an equal volume at a higher level. The pressure of the atmosphere, therefore, does not increase uniformly with the depth, as would be the case if the air were an ocean of incompressible fluid, but increases more rapidly than the depth. This becomes at once obvious by considering that any layer of the air near the earth not only has above it a number of overlying strata greater than any higher layer, but also that the mass of air in any stratum is greater than is the mass in any higher stratum. A given vertical height near the earth corresponds, therefore, to a greater change of pressure than the same vertical height at a greater altitude.

181. The Distinction between Liquids and Gases. The points in which liquids and gases resemble each other, and the laws applying in common to both of these bodies, are due, as has been seen, to the fact that the particles of each are free to move. Each of these bodies, however, has characteristics peculiar to itself; and there are accordingly a number of respects in which the two are unlike, and a number of laws that are true of one that do not apply to the other.

Among these points of difference may be mentioned the property of compressibility. It is found that liquids, when subjected to pressure, undergo only very slight changes of volume, while gases, under the same conditions, change their volume to a notable degree, depending upon the pressure applied. The volume of a gas, therefore, is not only a function of the mass of the gas, but also of the pressure to which it is subjected, as will be shown in § 182.

An interesting comparison may be drawn between the characteristics of the three states of matter, the solid, the liquid, and the gaseous, as regards the shape as well as the volume of each. A solid body, because of the rigidity of its particles, possesses volume and also shape; a liquid, since it possesses a definite surface, and adapts itself to the shape of the containing vessel, has volume but no shape; while a gas, inasmuch as it expands and fills the entire available space, whatever its shape or size may be, has neither volume nor shape. A further very weighty distinction between liquids and gases lies in the fact that a liquid has a free surface and can form into drops, while a gas has no free surface and cannot form drops.

182. The Volume of a Gas is Inversely Proportional to the Pressure—Boyle's Law. While Galileo and his pupil Torricelli in Italy were investigating the pressure of the air, and while Otto von Guericke, in Magdeburg, was studying the subject with the aid of the air-pump, two other investigators in other parts of the world were independently examining the relation that exists between the volume of a gas and its pressure. As a result of their experiments, Robert Boyle in England, and Mariotte in France, announced the law that is still known by the names of these discoverers, namely, *the temperature remaining constant, the volume of a gas varies inversely as the pressure to which it is subjected.* Thus if V represents the volume of a mass of gas at the pressure P, and if V' represents the volume of the same mass of gas when the pressure becomes P', it follows from Boyle's law, or Mariotte's law, that

$$V : V' = \frac{1}{P} : \frac{1}{P'},$$

or

$$\frac{V}{P'} = \frac{V'}{P},$$

from which

$$P\,V = P'\,V'$$

That is, the product of the volume of a mass of gas and the pressure to which it is subjected is equal to the product of the new volume of the same mass of gas and the pressure to which it is then subjected. In other words, *the temperature remaining the same, the product of the volume and the pressure of a gas is constant.* This is the form in which Boyle's law is usually stated.

183. The Density of a Gas is Directly Proportional to the Pressure. When the volume of a mass of gas is made to vary, that is, when a mass of gas is compressed into a smaller volume. or expands to fill a larger one, the density (p. 77) changes at the same time, becoming greater as the volume becomes smaller, and smaller as the volume becomes greater. The density of a gas accordingly, varies inversely as the volume. If d, therefore, represents the density of a mass of gas when its volume is V, and d' the density of the same mass of gas when its volume has been changed to V', P and P' being the pressures in the two cases respectively,

$$d : d' = \frac{1}{V} \quad \frac{1}{V'},$$

or

$$\frac{d}{V'} = \frac{d'}{V},$$

from which

$$\frac{d}{d'} = \frac{V'}{V}.$$

But by Boyle's law,

$$\frac{P}{P'} = \frac{V'}{V},$$

and therefore,

$$\frac{d}{d'} = \frac{P}{P'}.$$

From this it is seen that *the density of a gas is directly proportional to the pressure to which it is subjected.*

184. Proof of Boyle's Law for Pressures Greater than One Atmosphere. The device for proving Boyle's law consists of a bent glass tube, Fig. 121, the arms being of unequal length, the shorter arm closed, the longer arm open to the air. Mercury is then poured into the tube until it fills the bend of the tube, and thus imprisons a quantity of air in the shorter arm. When the mercury stands at the same level in both arms, the pressure of the air enclosed in the shorter arm is that of one atmosphere, since the surface of the mercury in the longer, open arm is directly exposed to the atmospheric pressure, and the pressure must be equal at all points of the same level. The pressure of the enclosed gas is thus known to be one atmosphere, and its volume is determined from the attached scale.

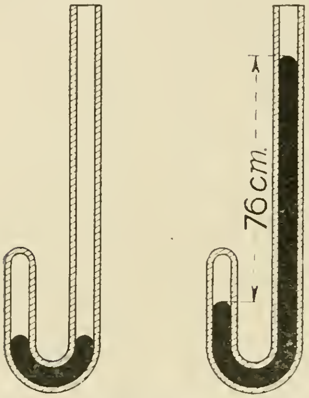

FIG. 121. Proof of Boyle's law for pressures greater than one atmosphere.

If mercury is now poured into the longer arm, the pressure upon the enclosed mass of gas increases and its volume diminishes. When the height of the mercury column in the longer arm is 76 centimeters above the level of the mercury in the shorter arm, the pressure upon the enclosed mass of gas has become equal to two atmospheres, that is, the pressure upon it has been doubled, and its volume is then observed to be reduced to one-half. It is thus seen, certainly for pressures greater than one atmosphere, that the product of the pressure and the volume of a gas remains constant.

185. Proof of Boyle's Law for Pressures Less than One Atmosphere. The apparatus for proving Boyle's law for pressures less than one atmosphere consists of a long, deep cistern, Fig. 122, and a glass tube closed at one end, partly filled with mercury and partly filled with the gas to be investigated. If the open end of the tube is placed in the cistern of mercury, the tube may be lowered until the level of the mercury within the tube is the same as the level of the mercury in the cistern, and in this position, the gas above the mercury in the enclosed tube is at atmospheric pressure. The volume of this gas at atmospheric pressure is determined, as in the previous experiment, by an attached scale.

If the tube is now raised, the mercury it contains is no longer at the level of the mercury in the cistern, but stands above the latter surface, and the gas in the upper part of the tube expands. In all positions the pressure within the tube, at the level of the mercury in the cistern consists of two parts, namely, the weight of the column of mercury *AB* above this level, and the pressure of the gas in the upper part of the tube. That is, the pressure of the air upon the surface of the mercury in the cistern sustains the weight of a certain column of mercury, plus the pressure of the gas above this column. The pressure of the gas enclosed within the tube is therefore always less than one atmosphere, and, indeed, it is always equal to the pressure of one atmosphere minus the weight of the column of mercury below it (p. 191).

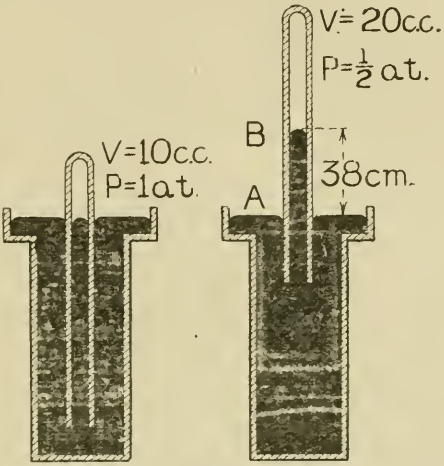

FIG. 122. Proof of Boyle's law for pressures less than one atmosphere.

If the tube is raised to such a height that the column of mercury *AB* is 38 centimeters, it will be found that the gas above the mercury has doubled its original volume. Since this column of mercury is only 38 centimeters high, the pressure due to its weight is only half an atmosphere, and accordingly it follows that the gas enclosed above the mercury column must also exert a pressure of 38 centimeters of mercury, or half an atmosphere, because the two together exert a pressure of 76 centimeters of mercury, or an entire atmosphere. While, therefore, the volume of this gas has been doubled, its pressure has been reduced one-half, so that, in this instance also, that is, for pressures less than one atmosphere, the product of the pressure and the volume is constant.

186. Boyle's Law Not Strictly True. In discussing Boyle's law, it should be mentioned that the above law holds true only when gases are far removed from their points of liquefaction. All gases, in fact, show very slight deviations from Boyle's law, and these deviations become greater as the pressure is approached at which a gas changes from the gaseous to the liquid state. It is then found that for all gases except hydrogen the volume diminishes more rapidly than would be indicated by Boyle's law. With hydrogen at ordinary temperatures, however, the volume does not diminish quite as rapidly as the pressure increases.

187. The Manometer. A manometer is any device for measuring pressures. Among the many existing forms, mention may be made of three that depend for their action upon laws of liquid and gaseous pressures. These are the open-arm manometer, the closed-arm manometer, and the barometer.

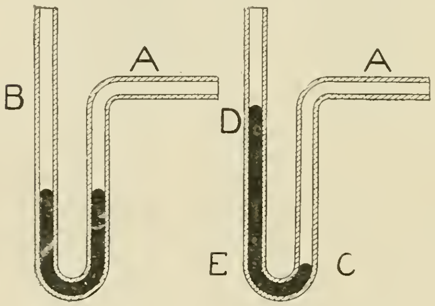

FIG. 123. The open-arm manometer.

The open-arm manometer consists of a bent tube, Fig. 123, partly filled with mercury, the end *A* being connected with the gas the pressure of which is to be measured, while the end *B* is open to the air. When the pressure in *A* is equal to one atmosphere, the mercury stands at the same level in both arms. When the pressure in *A* becomes greater than one atmosphere, the mercury falls on this side of the bent tube and rises on the other, as at *C* and *D*. The pressure at *C*, and therefore also in *A*, is then greater than the pressure of one atmosphere by the weight of the column of mercury *DE*, and this may be read at once from an attached scale.

The closed-arm manometer, Fig. 124, consists of a bent tube as before, but in this case the arm *B* is closed. The lower part of the tube is then filled with mercury, and air or some other gas is thus confined above the mercury in the closed arm. The level of the mercury is then adjusted so that it stands at the same height in the two arms when the pressure in *A* is one atmosphere.

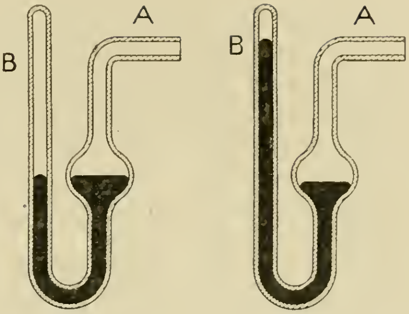

FIG. 124. The closed-arm manometer.

If the pressure in *A* becomes greater than one atmosphere, the mercury rises in the closed arm, and the volume of the gas in this arm becomes smaller under the increased pressure. Therefore, from the decrease in the volume of this gas, the pressure to which it is subjected may be calculated by Boyle's law, and thus the pressure in *A* may be determined.

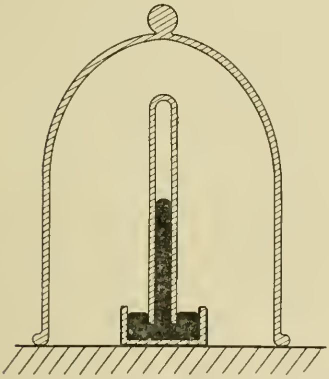

FIG. 125. The barometer used as a manometer.

The barometer, Fig. 125, is often a most convenient form of manometer, and is useful in measuring the pressure of the gas or air in the region where it is placed, since the height of the mercury within the tube is directly proportional to the pressure upon the surface of the mercury in the cistern.

188. The Siphon. The siphon is a device for transferring liquids from an upper to a lower vessel, and depends in its action upon the pressure of the air. The apparatus consists of a tube open at both ends and so bent that one arm is longer than the other. When the tube is filled with liquid and inverted, the shorter arm being placed beneath the surface of the liquid in the upper vessel, the liquid will flow from the longer arm. Moreover, this flow will continue, the liquid rising in the shorter arm and passing over the bend of the tube into the longer arm, until the level of the liquid in the upper vessel falls below the open end of the shorter tube. At first, this phenomenon seems to contradict the principle that a liquid cannot rise to a height greater than its source; but a consideration of the pressure of the air and the height of the column of liquid this pressure supports, leads at once to the following evident explanation.

We have learned that the pressure of the air sustains a column of liquid of such height that the pressure due to the weight of this liquid is equal to the atmospheric pressure. Moreover, the pressure at any point within such a column is the weight of the column of liquid above that point, or, in other words, the pressure at any point within a liquid column is the pressure of one atmosphere minus the weight of the column of liquid beneath it.

Application is made of these principles in explaining the action of the siphon. If *B*, Fig. 126, is the highest point of the siphon, the pressure of the air at *A* upon the surface of the liquid in the upper vessel supports the column of liquid *AB*, and the pressure of the air at *C* upon the open end of the longer tube sustains the column of liquid *CB*. If we consider a particle of the liquid at *B* and determine the pressures upon it, we shall find that this particle is not in equilibrium, but that the pressure in one direction is greater than that in the opposite direction. The pressure urging the particle to the right is the pressure of one atmosphere minus the weight of the column of liquid beneath it, that is, minus the weight of the column of liquid the height of which is *a*. The pressure at *B* urging the particle to the left is the pressure of one atmosphere minus the weight of the column of liquid the height of which is *b*. The pressure, therefore, upon a particle of liquid at *B*, urging it to the right, is greater than the pressure upon the same particle urging it to the left by an amount equal to the weight of a column of liquid, the height of which is

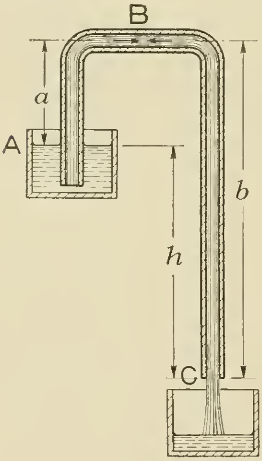

FIG. 126. The siphon.

$$b - a = h.$$

Accordingly, the liquid flows from the longer arm of the siphon due to an unbalanced pressure equal to the weight of a column of the liquid, the height of which is the difference between the lengths of the two arms. The flow will continue from the longer arm, until the level of the liquid in the upper reservoir falls below the end of the shorter tube, when air will enter instead of the liquid, and the action of the siphon will cease.

189. The Sprengel Pump. A form of air-pump has been devised, known as the Sprengel pump, in which advantage is taken of the Torricellian vacuum above the column of mercury in the barometer tube. For this purpose, the tube *AB*, Fig. 127, should be considerably longer than 76 centimeters, and should have an enlargement *C* at its upper end. Into this enlargement projects the tube E which connects with a reservoir of mercury *F*, so that mercury may enter the enlargement one drop at a time and pass down the tube *BA* drop by drop. As these drops of mercury enter the tube *BA*, they imprison little columns of the air that has entered the tube from the enlargement *C*, and so push these little columns of air before them as they move down. The air is thus gradually removed from this enlargement; and the pressure of the air upon the surface of the mercury in the cistern at *A* causes the mercury ultimately to stand at the height *AB* in the barometer tube. When the height of this mercury column approaches 76 centimeters, the vacuum in *C* is very nearly perfect.

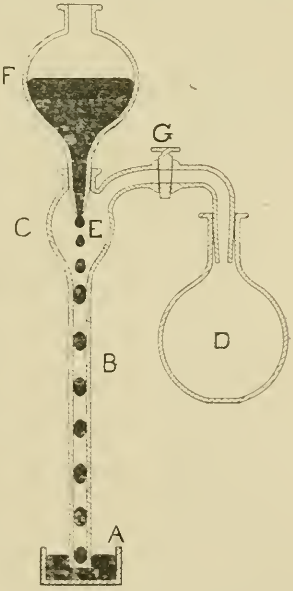

FIG. 127. The Sprengel pump.

The reservoir *D*, which is to be exhausted, may now be connected with the enlargement *C* by the stop-cock *G*, and the air in *D* expands and flows into *C* until the pressures in *D* and *C* are equal. The action of the pump then being continued, the drops of mercury, as they fall from the tube *E,* again imprison little columns of air, and serve as pistons to push these columns of air down the tube. The air thus carried down bubbles out through the mercury in the cistern at *A,* and with each little bubble there is thus removed a portion of the air originally in *D.* By continuing this process, the air in the reservoir *D* may be surely, though slowly, carried to any required degree of exhaustion.

190. The Fountain in Vacuo. The experiment known as the *fountain in vacuo* depends upon the principle of the Sprengel pump. The flask, or receiver, *A*, Fig. 128, is closed by a cork, through which project two tubes, a short one *BC,* dipping into a vessel of water at *C*, and a longer one *DE,* terminating in a cistern of water at *E.* If the flask *A* contains originally a little water, this water now flows out through the tube *DE,* thus producing a partial vacuum in *A.* The pressure of the air upon the water at *C* then forces water through the tube *CB* into the flask *A.* This water in flowing out through the tube *DE,* imprisons little columns of air, which are carried down the tube and bubble out through the water in the cistern, as in the Sprengel pump. Therefore, as air is being continually removed from the receiver *A,* the vacuum here grows more nearly perfect, and the jet of water, rising like a fountain from *B,* strikes with greater and greater force against the top of the receiver.

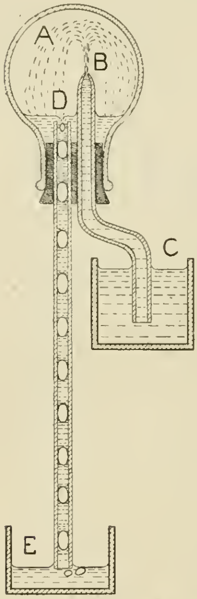

Fig. 128. Fountain in vacuo.

191. Mariotte's Bottle. We have seen, (p. 185), that when a liquid flows from an opening near the bottom of a reservoir, the flow takes place with a velocity that depends upon the depth of the orifice beneath the surface. As the flow continues, this depth becomes less and less, and the velocity with which the liquid issues is, therefore, a decreasing one. Mariotte devised a means for maintaining this flow perfectly constant, notwithstanding the fact that the level of the liquid in the reservoir continually falls, and the height of the column of liquid above the orifice, therefore, constantly diminishes.

This device, which is still known as Mariotte's bottle, Fig. 129, consists of a bottle, or flask, closed at the top, through the cork of which, however, passes a tube *AB*, open at both ends, and extending well down into the interior of the bottle. Near the bottom of this flask is an aperture *C*, through which the liquid is to flow, and this may be closed temporarily by a cork. If the flask is now filled with water to any convenient height, as *D*, the water will stand at this same level in the tube *AB*. If the cork at C is now withdrawn, water will flow from this orifice, and the level of the water in the tube *AB* will fall to the end of the tube *B*. As the flow from *C* continues, bubbles of air enter the water at *B* and rise to the surface at *D*, the air entering the bottle to take the place of the water that flows out. Under these conditions, *water flows from C with a perfectly constant velocity*, and will continue to do so until the level *D* of the water in the bottle falls below the end *B* of the tube. Since this flow is constant, it follows that the resultant pressure at *C* urging the water outward must also be constant, and it is, indeed, upon this constant resultant pressure at *C* that the action of the bottle depends.

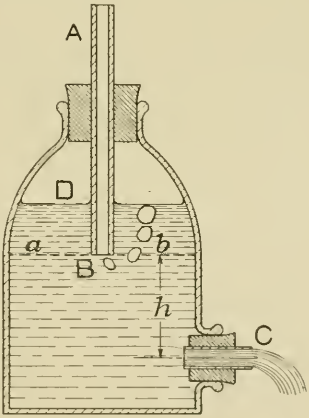

Fig. 129. Mariotte's bottle.

Inasmuch as air enters the bottle at *B*, the end of the tube *B* is at atmospheric pressure, and therefore all points at the same horizontal level, that is, all points in the plane *ab*, are also at atmospheric pressure. The pressure at *C* acting inward is simply the pressure of one atmosphere, since this is also open to the air. The pressure, however, at *C* acting outward is the pressure of one atmosphere, which exists at *ab*, plus the pressure due to the weight of the column of water, the height of which is the depth of the opening *C* below the plane *ab*. This height, which we may represent by h, remains perfectly constant, and is entirely independent of the distance of the surface D above the open end of the tube B. Accordingly, the water flows from C with a resultant constant pressure equal to the weight of a column of water the height of which is h. If the tube AB is pushed farther down into the bottle, the height h becomes less, and the velocity with which the water issues from the orifice is also less.

Mariotte's bottle, therefore, provides not only for a constant flow of water, but for a flow that remains constant with any velocity we may desire, depending only upon the height of the tube B above the opening C. If the tube AB is pushed down until the end B is at the same level as C, the distance h becomes zero, the resultant pressure is then also zero, and the flow ceases. Whatever may be the position of the tube AB, the pressure of the air in the upper part of the bottle is the pressure of one atmosphere minus the weight of the column of water DB.

192. Diminution of Pressure due to Velocity. A phenomenon of unusual interest, connected with the flow of fluids, is observed when a liquid or a gas issues from an orifice forming a jet. To illustrate, and at the same time explain this phenomenon, we may assume that a liquid, water for instance, flows through a tube that has a constriction, Fig. 130, so that one part of the tube is much smaller in cross-section than the remainder. a, a', a'' may thus represent the areas of these different cross-sections, a' being less than a or a''. If the water flows through the tube without rupture, that is, if the tube is completely filled at every point, the quantity of water passing any cross-section in a given time must equal the quantity passing any other cross-section in the same time. Accordingly, the velocity with which the water flows will be different in different parts of the tube, being least where the cross-section is greatest, and great-

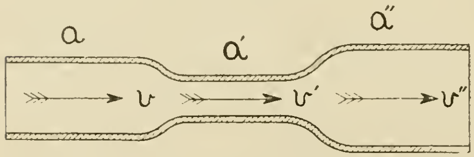

FIG. 130. Flow through a tube of variable cross-section.

est where the cross-section is least, that is, the velocity is everywhere inversely proportional to the cross-section. From this it follows that the product of the cross-section at any point and the velocity at that point is constant, or that

$$av = a'v' = a''v'',$$

where v, v', v'' are the velocities of the water at a, a', and a'' respectively.

As the water approaches the constriction, its velocity becomes greater, increasing from v to v', and therefore the momentum of a mass m of water increases from mv to mv'. But by Newton's second law of motion (p. 105), a force is necessary to produce a change of momentum, and accordingly a force, or, in other words, a pressure, exists in the water between a and a' urging it forward with higher speed, that is, increasing its velocity. This pressure is transmitted to the sides of the tube, and so the sloping walls of the tube are likewise under pressure from the liquid.

As the water passes the constriction with a high velocity, it then enters the larger part of the tube where it is compelled to move with a lower velocity. During the passage, therefore, from a' to a'', the momentum of the water must be decreased from mv' to mv'', and again a force is required to produce this change of momentum. Moreover, as the velocity is now being decreased this force, or pressure, must act in a direction opposite to that in which the water is flowing, so that the sloping walls on this side of the constriction must likewise sustain a pressure from the liquid equal to the force required to change the momentum of the water.

In the constriction, however, the water is flowing, by Newton's first law, with uniform motion, and therefore, neglecting friction, it is moving without the action of any force whatever, that is, it is not under pressure itself nor does it exert pressure upon the parallel sides of the tube. In other words, the pressure exerted before and after the constriction is entirely expended in changing the momentum of the water at these points, and therefore is not transmitted to the liquid in the constriction. At a' therefore, the pressure is less than that at a or a'', and so if a'' is open to the air, the pressure at *a′* is less than atmospheric pressure. Accordingly, if a small opening were made at *a′*, air would enter this opening and mingle with the water as it rushes past.

If a tube were inserted in the opening, Fig. 131, the lower end dipping beneath the surface of a cistern of water, the air in the upper part of this tube would become exhausted, and the water would rise in the tube above the level of the water in the cistern. It is interesting to observe, however, that if an opening were made in the sloping sides of the tube, as at *b*, or *c*, the water would be forced out. That is, at *b* or *c* the water is at a pressure greater than one atmosphere, while at *a′* the pressure is less than one atmosphere. The principle upon which this phenomenon depends may therefore be expressed by saying that when a liquid or a gas flows through a tube of varying cross-section, the pressure it exerts is least where the velocity is greatest and greatest where the velocity is least.

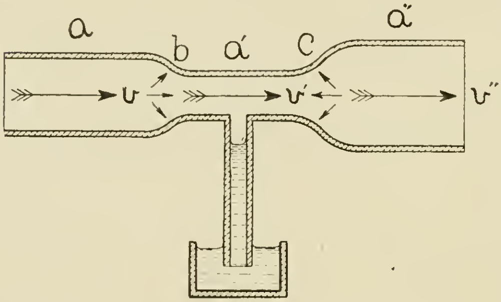

Fig. 131. Lateral diminution of pressure due to velocity.

193. The Atomizer. The operation of the device known as the atomizer depends upon the principle here set forth that there is a lateral diminution of pressure due to the velocity of an issuing jet. In its essential features, the atomizer consists of a horizontal tube *AB*, Fig. 132, contracted to a fine opening at *B*, and so supported that the opening at *B* is directly above the tube *CD*. This tube, which is likewise contracted to a small opening at *D*, dips beneath the surface of the liquid in the cistern or reservoir at *C*. When air is forced through the tube *AB*, it issues from *B* as a fine jet, moving with a high velocity; but it soon expands into the air beyond, and thus has its velocity diminished. The air is under pressure in the tube *AB*, where its velocity is being increased, and is also under pressure when it expands into the atmosphere beyond *D,* where its velocity is being diminished. At *B,* therefore, where the velocity is greatest, there is a lateral diminution of pressure all around the issuing jet, and the air immediately about this jet is at less than atmospheric pressure. In the upper part of the tube *CD* the air accordingly expands, passing out through the orifice at *D* into this region of lower pressure. Since this leaves a partial vacuum in the tube *CD*, the pressure of the air upon the surface of the liquid in the cistern at *C* causes the liquid to rise and fill this tube. A tiny drop of liquid thus gathers at the point of the tube *D,* and is blown into spray by the jet of air issuing from *B.* Its place, however, is immediately taken by another drop, due to the pressure of the air at *C,* and this drop is likewise, minutely divided and scattered by the impact of the jet of air, and so the action of the instrument is continuous.

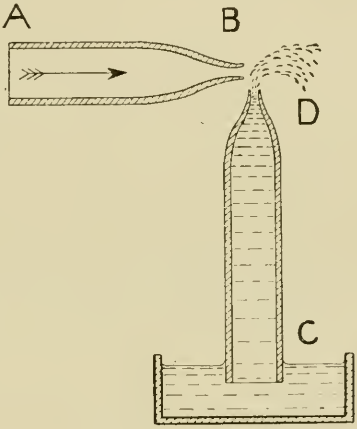

FIG. 132. The atomizer.

An important application of this principle is made in the use of the injector, by means of which water is introduced into the boiler of a steam engine. Steam from the boiler passes through a long tube, from which it issues as a fine jet and passes again into the boiler. The region all about the jet is connected with the supply of water, and the extremely high velocity of the issuing jet of steam produces so great a lateral diminution of pressure, that water enters the jet upon the principle of the atomizer, is blown into spray, and is thus carried in the form of minute drops with the steam into the boiler.

CHAPTER XIII.

SURFACE TENSION AND CAPILLARITY.

194. Forces of Adhesion and Cohesion. In the foregoing treatment of fluids and fluid pressures, no consideration has been given to the attractions, or forces, exerted by the particles of a fluid upon one another or upon the walls of the containing vessel. Our attention now, however, must be directed to a class of phenomena, observed in liquids but not at all in gases, which have their origin in the attraction of the ultimate particles of matter for one another.

That the molecules of a liquid exert an attractive force upon one another and also upon the molecules of a solid body may be seen by withdrawing a glass rod from a vessel of water. The water remaining upon the rod gathers into a drop at the bottom, Fig. 133, notwithstanding the fact that a force equal to its weight tends to separate it from the rod. When, indeed, the separation is finally effected and the drop falls, it is found that the rupture has taken place not between the water and the rod, but between the particles of water themselves, for the end of the rod is still wet. It is thus seen that the force of cohesion of the water over an area equal to the neck of the drop is equal to the weight of this much water. Moreover, since particles of water still cling to the rod, it is obvious that the adhesive force between the glass and the water is greater than the cohesive force of the water.

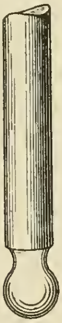

FIG. 133. Adhesion and cohesion of water.

A conception of the magnitude of these adhesive and cohesive molecular forces may be obtained by suspending a perfectly clean glass plate from one side of a balance, and counterpoising it by weights upon the opposite side. If the glass plate is now allowed to touch the surface of water in a vessel placed beneath, it will be found that the balance is no longer in equilibrium, but that the glass is held upon the water surface by the force of attraction between the two, Fig. 134. Additional weights may now be placed upon the opposite side of the balance, and the force necessary to pull the plate away from the water may thus be determined. When this has been accomplished, it is found that the glass has not been parted from the water, since the under surface of the plate is still wet, but that the force employed has been expended in separating the particles of water from one another, or, in other words, in overcoming the cohesion of the water. Here again we have proof that the adhesion between glass and water is greater than the cohesion of water.

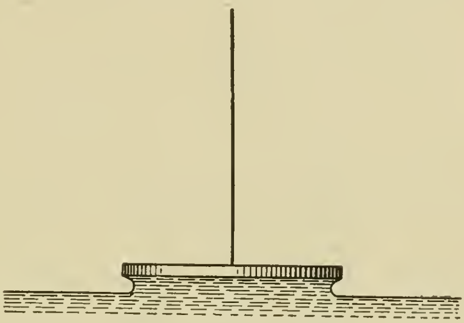
FIG. 134. Measuring the cohesion of water.

If the experiment is now repeated, a vessel of mercury being substituted for the water, it will be found that the glass and the mercury likewise attract each other, and that the force required to separate them is between three and four times as great as that required to remove the glass from the water. Moreover, it will now be seen that the glass is not wet by the mercury, since no particles of mercury cling to the under surface of the plate. The force, therefore, required to separate the plate from the mercury is a true measure of the attraction, or adhesion, of these two substances for each other; and although this force of adhesion is considerable, it is nevertheless not so great as is the force of cohesion between the particles of mercury.

In general it may be remarked that if a solid is wet by a liquid, the adhesion between the two is greater than the cohesion of the liquid; but if a solid is not wet by a liquid, the adhesion between them is less than the cohesion of the liquid. The adhesive force between liquids and solids varies greatly with the nature of these substances, as may be seen from the fact that gold is readily wet by mercury, while glass is not.

195. Change of Surface Form Near a Wall. Since the particles of a solid exert this great attractive force upon the particles of a liquid, it is not surprising to find that the form of a liquid surface changes near the walls of the containing vessel, becoming concave or convex according as the adhesion between the walls and the liquid is greater or less than the cohesion of the liquid. This may be seen by considering the forces that act upon a liquid particle at the boundary between a liquid surface and a solid. Thus, in Fig. 135, let *AB* represent a liquid surface, and *CD* the wall of the containing vessel, or the wall of a solid body partly immersed in the liquid. The particles of liquid in the quadrant *BAD* exert attractive forces upon the liquid particle *A*, and these forces may be combined into a resultant force *P*, making an angle of 45° with the surface of the liquid. The attractions upon *A* of the particles of the solid wall may likewise be combined into a resultant force *Q*, which is directed towards the wall and normal to it. Besides these forces, there is the weight of the particle *A* which acts vertically downward. The force *P* may be resolved into two components, one, *P* cos 45°, or *P′* acting in the horizontal direction away from the wall, and the other, *P* sin 45°, acting in the vertical direction downward. The latter component may be added to the weight of the particle, thus forming a total vertical downward force of $mg + P \sin 45°$. If we now choose one of the horizontal directions, *AE*, for instance, as positive, we find that the particle *A* is acted upon by a resultant horizontal force equal to *Q*—*P′* and by the resultant vertical downward force just mentioned.

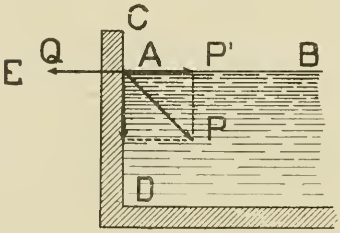

FIG. 135. Forces of attraction near a wall.

There are now three possible conditions of equilibrium depending upon the relative magnitudes of *Q* and *P′*.

1. If $Q > P'$, the resultant horizontal force is directed from *A* toward the wall, Fig. 136, and this may be compounded with the total vertical force acting on *A* to give a final resultant *AR* which lies in the quadrant *EAD*. We have learned, however, (p.

172) that the surface of a liquid must be perpendicular to the resultant of all the forces acting, hence at A it cannot be horizontal. The liquid at this point must therefore rise and be inclined to the wall at such an angle that the surface is perpendicular to the resultant AR. As the distance from the wall increases, the attractive force due to the wall becomes less, the resultant force acting on the particle approaches more nearly the vertical, and the surface soon becomes horizontal. Therefore, if the adhesion between a solid and a liquid is greater than the cohesion of the liquid, that is, if the solid is wet by the liquid, the latter will rise upon the side of the solid, and the surface will be concave upward.

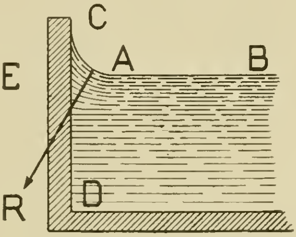

FIG. 136. Elevation of surface when the wall is wet.

2. If $Q < P'$, the resultant horizontal force upon A is directed away from the wall, Fig. 137, and this, combined with the total vertical force, yields a resultant which lies in the quadrant BAD. Again the surface of the liquid must be perpendicular to this resultant, hence, in this case also, the surface near the wall is not horizontal, but is rounded away from the wall, the inclination at each point being such that the surface is everywhere perpendicular to the resultant AR. The total resultant force acting on a surface particle rapidly approaches the vertical direction as the distance from the wall increases, so that the surface soon becomes horizontal. Accordingly, if the cohesion of the liquid is greater than the adhesion between the liquid and the solid, or, in other words, if the liquid does not wet the solid, the liquid is depressed around the edges, and the surface is convex upward.

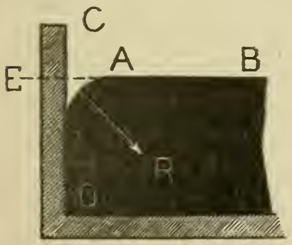

FIG. 137. Depression of surface when the wall is not wet.

3. If $Q = P'$, the resultant horizontal force is zero, the only force then acting upon the particle is the total resultant vertical force, and the surface is horizontal throughout. In practice,

however, it seldom happens that the cohesive force of a liquid is thus exactly equal to the adhesive force between the liquid and an immersed solid.

196. The Surface Film and the Normal Pressure. Before investigating further the forces of attraction between a liquid and a solid, it is necessary to consider in detail the unusual molecular condition that exists at the surface of a liquid. Let *m,* Fig. 138, represent a liquid molecule, surrounded by an imaginary sphere, which we may call its sphere of activity, or sphere of attraction, (p. 6), and let this molecule be situated so far beneath the surface of the liquid that the sphere of attraction does not touch the surface. In this case the molecule is surrounded by similar molecules on all sides, so that the attractions of neighboring molecules are perfectly symmetrical.

If, however, the molecule is situated exactly in the surface, as at *m''*, this molecule is attracted laterally and downward by the adjacent molecules, but can experience no attraction upward, since half of its sphere of attraction is above the surface. The attraction of the molecules in the lower hemisphere in any direction may be resolved into a horizontal and a vertical component. The horizontal components, by symmetry, hold one another everywhere in equilibrium, while the vertical components combine into a single force acting at right angles to the surface. All surface particles, therefore, sustain a force which acts normally to the surface, and is not unlike a pressure.

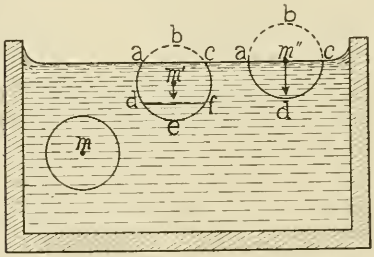

FIG. 138. The normal molecular pressure upon the surface of a liquid.

If the particle is situated, as at *m'*, so that a portion only of its sphere of attraction projects above the surface, a plane *df,* symmetrical with the surface, may be laid through this sphere, cutting off the portion *def,* similar in every way to the portion *abc.* The attractions of the molecules within the portion *adfc* are everywhere symmetrical and therefore need not be considered. It follows, therefore, that the attractions of [illegible]cules in the spherical segment *dcf* are alone effective. [illegible]se combine into a resultant force acting vertically downward.

What is true of these molecules is also true of all particles lying nearer the surface than the range of their molecular attraction, for these do not have their spheres of attraction completely filled with liquid. If we imagine a plane laid within the liquid and parallel to the surface, the distance of this plane from the surface being the radius of the spheres of attraction, we see that the molecular relations over this superficial layer are different from the conditions existing in the main body of the liquid. The particles of this layer are under the influence of molecular forces acting everywhere perpendicular to the surface, the effect of which is to subject the entire superficial layer to a normal pressure.

197. The Energy of the Superficial Film. Since this normal pressure exists only at the surface of a liquid, it follows that work must be done in transferring a particle from the interior, where this pressure does not exist, up into the surface, against the action of this normal molecular force. Therefore, any change, in the form of a liquid mass that brings a larger number of particles into the surface, that is, any change that increases the area of the surface, such as flattening out a spherical drop into a film, Fig. 139, requires the expenditure of work. From this it follows that a liquid surface represents a certain amount of work done; and the work that has thus been performed in causing the surface to be increased in extent remains in the superficial film in the form of potential energy.

Accordingly, the surface of a liquid represents a definite store of potential energy; and this amount of energy will re-

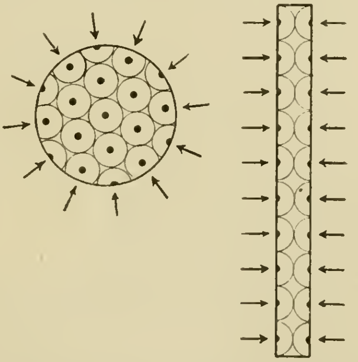

FIG. 139. Work must be done in bringing liquid particles from the interior into the surface.

main undiminished in the bounding film so long as the area of the liquid surface undergoes no change. Because of the universal principle, however, that potential energy tends towards a minimum (p. 122), the energy of the film, which is proportional to the extent of the surface, will, if possible, decrease in amount. But the store of energy, thus represented by the surface film, can become less only as this film itself becomes smaller. Every liquid surface, therefore, is under a strain, or tension, always tending to contract, and always tending to reduce its area to the smallest possible value.

198. Surface Tension. Although the contraction of a liquid film is due, as we have seen, to the inherent tendency of potential energy to decrease in amount, the effect thereby produced is the same *as if* forces of attraction existed between the molecules, acting everywhere parallel to the surface, and causing the superficial layer to contract. In the year 1805, Thomas Young referred to the surface of a liquid as being under tension; and since that time the expression ''surface tension'' has been used to denote this peculiar contractile condition of the surface. Indeed, the superficial layer of a mass of liquid has been likened to a thin, elastic skin, or membrane, tightly stretched over the surface, always under tension, and consequently always tending to contract.

While no such skin, or membrane, actually covers a liquid, and while, so far as is known, no forces other than those of cohesion really exist between the molecules of the surface, the analogy suggested by Thomas Young has been adopted, and the contraction of a liquid surface is now universally said to be due to its surface tension. This, however, is only figurative; and the true cause of the effect which this expression represents is found in the constant tendency of the potential energy in a free liquid surface to assume a minimum value.

199. Spherical Form of Drops Due to Surface Tension. From the above consideration it is evident that a mass of liquid, if free to contract under the action of forces resident in its surface, will be forced into the form having the smallest possible superficial area. Since a sphere has the minimum surface for a given volume, it follows that a mass of liquid, if uninfluenced by other forces, assumes the spherical form, as may be seen in drops of rain and drops of dew. Indeed, the compression of a body of liquid into the spherical form may be regarded as occasioned by the contraction of this skin, or membrane, with which the surface of a liquid may be conceived as covered. It is, moreover, because of this elastic, tightly-stretched, superficial layer, which strongly suggests a membrane under tension, that phenomena which depend upon this surface film are included under the general head of surface tension.

The oscillations of a falling drop of water (p. 187) are due to the superficial contraction of the elongated drop; but the inertia of its particles carries it past the spherical form until it is elongated in the opposite direction, and the same process is then repeated. When a jet of water falls vertically downward, the constrictions that form in the cylindrical part of the jet just before the drop breaks away are due to the contraction of the superficial film. The more complete theory of surface tension, which cannot well be taken up here, shows that a cylinder of liquid is in equilibrium with its bounding film so long as its length is not greater than its circumference. When, however, this limit is exceeded, the contractile tendency of this film causes a constriction, or neck, to form in the cylinder, and the rupture is effected at this point. It is interesting to observe that after the drop has fallen, the superficial contraction of the liquid forming the neck forces the latter also into the spherical form, so that a tiny droplet appears between every two drops of the falling stream, as shown in Fig. 118.

Another beautiful illustration of the fact that a cylinder of liquid develops necks, whenever its length is greater than its circumference, is found in the web of the common geometrical garden spider. When examined with a lens, each strand of this web is found to be covered with minute globules of a sticky substance, serviceable in catching insects, the whole forming the appearance of tiny pearl beads threaded upon a string with great regularity. These globules are not put in position after the web is made, for this would be quite impossible, as their number often exceeds a million. To accomplish this distribution, the spider, as it spins its web, spins also around each strand, and concentric with it, a cylinder of this sticky material. Along this cylinder, constrictions, or necks, then develop at distances apart equal to its circumference, and the surface tension of the liquid soon forces the segments thus produced into the spherical form. Even the small amount of this substance still remaining upon the moistened thread gathers likewise into minute globules midway between the larger drops, as may be seen in Fig. 140, which is from a photograph of a spider's web taken under the microscope.

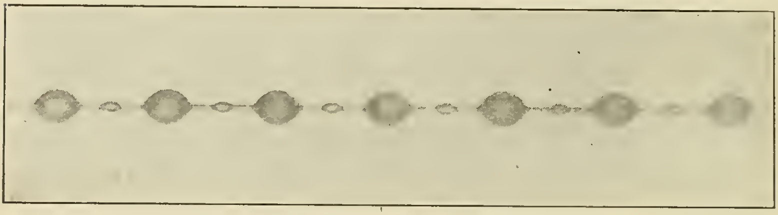

FIG. 140. A spider's web, showing globules of sticky material; magnified 250 diameters.

200. Measurement of Surface Tension and Surface Energy. An illustration of the tendency of a film to contract, and at the same time a measure of the surface tension of this film, may be obtained by withdrawing from a soap solution a ∧-shaped frame *AB*, Fig. 141, upon which is laid a freely sliding wire *CD*. A liquid film then occupies the area between the frame and the wire; and as this tends to contract, it draws the wire towards the upper part of the frame, thereby lifting it against the force of

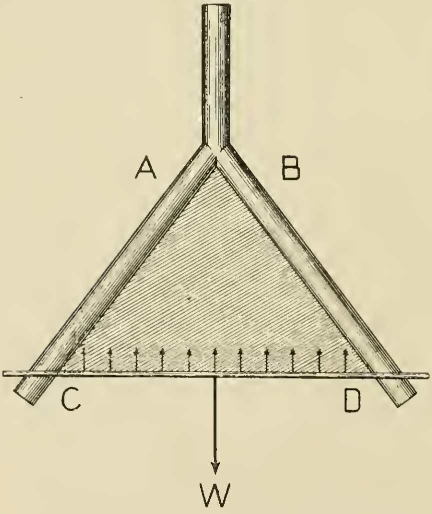

FIG. 141. Measurement of surface tension.

gravity. If T represents the tension existing upon a line of particles of unit length in each surface of the film, the total force applied to the movable wire is $2T \cdot CD$, since the film has two surfaces. The wire remains in equilibrium when the total upward tension upon it is equal to its weight mg, that is, when

$$2\,T \cdot C\,D = mg,$$

or

$$T = \frac{m\,g}{2\,C\,D} \text{ dynes per centimeter.}$$

If the frame has parallel arms, as in Fig. 142, the wire CD is lifted through the height CA, and each surface of the film in contracting has done an amount of work equal to

$$T \cdot CD \cdot CA.$$

The surface, however, has now decreased in extent by the area $CD \cdot CA$; and if E represents the energy per unit of area of the film, the total energy of the surface has been decreased by

$$E \cdot CD \cdot CA$$

But the potential energy the film has lost must equal the work the film has performed, hence

$$E \cdot CD \cdot CA = T \cdot CD \cdot CA$$

or

$$E = T.$$

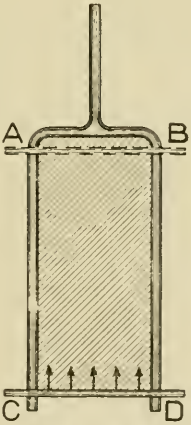

FIG. 142. Measurement of surface energy.

In other words, the energy of the film per unit of area is numerically equal to the tension of the film per unit of length.

To

201. Phenomena Due to Surface Tension. Numerous familiar illustrations give abundant proof of the existence of this membrane-like film over the surface of a liquid. Many little insects move about upon the surface of water, being supported solely by this film. The tiny indentation of each foot upon the surface shows that the weight of the insect exerts a pressure upon this membrane, but this is not sufficient to break through the film. If a common sewing needle is carefully laid upon the surface of water, Fig. 143, it will float upon the surface, provided only that care is taken to have the needle perfectly dry. As soon, however as it is even slightly moistened by the water, it is then beneath the film which would otherwise support it, and it falls to the bottom of the vessel.

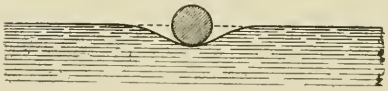

FIG. 143. A sewing needle floating upon the surface film of water.

If the wires of a small sieve are very fine, and are coated with a film of wax or paraffine so that they will not be wet by water, the sieve may be floated upon a water surface without sinking, the weight of the sieve being insufficient to break through the surface film upon which it rests. Water may also be carried in such a sieve, the weight of the water in this case being less than the pressure required to pierce the membrane-like film that forms over each small mesh.

When a camel's-hair brush has been wet with water, the hairs cling tightly together, not because the brush is wet, since the hairs stand apart when the brush is beneath the surface of the water, but because, when the brush is withdrawn, the contraction of the surface film draws the hairs together. Water, when poured from a beaker, may be prevented from running down the side of the vessel by holding a glass rod against the edge, and in this case the water follows the rod. When the water wets the rod, the liquid film about the moistened rod forms, as it were, an elastic tube through which the mass of the liquid is then poured.

The contractile tendency of this superficial film may be beautifully shown by dipping into a soap solution a wire ring two or three inches in diameter, provided with a handle. When this is withdrawn, a film of the solution, consisting almost exclusively of this superficial layer, covers the ring. If a loop of thread, previously moistened with the soap solution, is now thrown upon this film, Fig. 144, the loop will assume an irregular form, and will remain in equilibrium, since the soap film extends over the interior as well as the exterior of the loop, and the tensions within and without are the same. When the film within the loop is broken by touching it with a hot wire, the tension there disappears with the film, and the loop is immediately pulled out into a circular form by the now unbalanced tension of the exterior parts of the film.

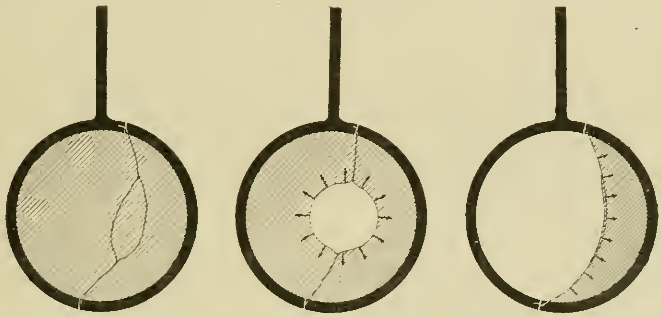

FIG. 144. The contraction of a soap film.

A conception of the very great surface tension of pure water, and of the amount by which dissolved substances weaken this tension, may be obtained by scraping a few flakes of camphor upon a perfectly clean surface of water. The camphor is slightly soluble in water, and dissolves more readily from the points and angles of the flakes than from the rounded surfaces and edges. Accordingly, the larger amount which thus dissolves from the points and corners weakens the surface tension to a greater extent than does the smaller amount which dissolves from the more rounded portions. The stronger surface tension where the solution has been less then overcomes the weaker tension where the solution has been greater, and the camphor flake is thus pulled about the surface by this unequal tension on its opposite sides. If a number of flakes are thus scraped upon a water surface, and if the water is perfectly clean and free from oily matter, the most curious and interesting rotary motions of the flakes are produced.

202. Surface Tension Over Curved Surfaces—Capillarity. A remarkable effect of the superficial tension of a liquid is the change in the normal pressure produced by this tension when the surface of the liquid is curved. In this case, the tangential forces *T,* as they act upon adjacent particles, make an angle with one another; and since these forces are then no longer equal and opposite, they yield a resultant *R*. This resultant, which always acts toward the center of curvature, is directed towards the liquid mass if the surface is convex, Fig. 145, and away from the liquid mass if the surface is concave Fig. 146. Accordingly, a convex surface sustains a normal pressure greater than that upon a plane surface, while a concave surface sustains a normal pressure less than that upon a plane surface. Moreover, as this normal component of the tension increases with the curvature, the change in the normal pressure is greater the smaller the radius of curvature of the surface.

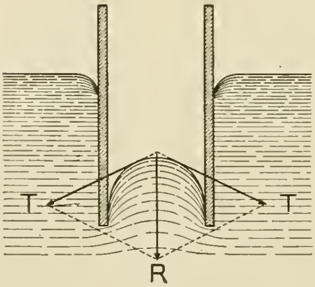

Fig. 145. The increased normal pressure upon a convex surface.

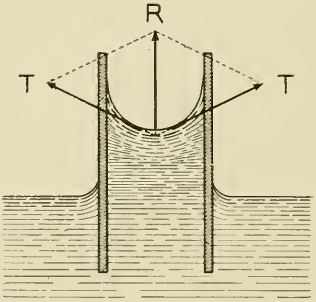

Fig. 146. The decreased normal pressure upon a concave surface.

The fact that the normal pressure due to this surface film is less over a concave surface and greater over a convex surface leads to most interesting results when a fine, or capillary, tube is immersed in a liquid. If the liquid wets the tube, as is the case with glass and water, Fig. 147, the surface of the liquid becomes concave upward, as we have seen (p. 211), and the normal molecular pressure is then less over this concave surface than over the plane surface. Accordingly, the liquid will rise in the tube and stand at such a height that the weight of this column, together with the decreased normal pressure upon the concave surface just equals the normal molecular pressure upon the plane surface of the liquid outside the tube. If the liquid does not wet the tube, that is, if the two behave like glass and mercury, Fig. 148, the surface of the liquid is everywhere convex and the liquid within the tube is depressed. Moreover, the amount of this depression is such that the increased pressure upon the convex surface of the liquid within the tube is equal to the weight of a column of the liquid the height of which is the depth of the depressed surface below the level of the liquid in the vessel. Thus it is seen that the phenomena of capillarity depend upon those of surface tension, and upon the changes in the normal pressure which this tension brings about when the surface of the liquid is convex or concave.

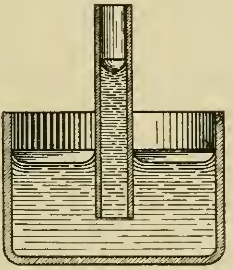

FIG. 147. Rise of water in a capillary tube.

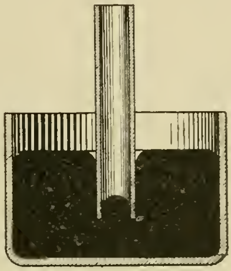

FIG. 148. Depression of mercury in a capillary tube.

203. Measurement of Surface Tension by Capillarity. This change in the normal pressure over a convex or a concave surface, enables us to regard the elevated column of liquid in a capillary tube as hanging by the superficial film, or membrane, from its point of attachment where liquid and wall meet. Accordingly, the tension per unit length at this point of juncture, multiplied by the number of units of length around the inner circumference of the tube, equals the weight of the elevated column of liquid. Accordingly, if T is the tension per unit length, r the radius of the tube, h the height of the elevated column, d the density of the liquid, and g the acceleration of gravity,

$$T \cdot 2\pi r = \pi r^2 h d g,$$

or

$$T = \frac{r h d g}{2}.$$

In a similar way, the elevation of the liquid upon the outside of the tube or against the wall of the containing vessel may likewise be regarded as a condition of equilibrium between the superficial tension and the weight of the elevated liquid. Here again the liquid hangs by its superficial, membrane-like film from the place where this is attached by adhesion to the wall of the vessel.

204. Capillary Phenomena. A porous body may be regarded as a collection of extremely irregular, minute capillary tubes, so that many familiar processes, such as drying the hands with a towel, the absorbent action of blotting paper, and the elevation of oil in the wick of a lamp, may, like other capillary phenomena, be referred to causes that have their origin in surface tension.

The above expression for the surface tension shows that the height to which a liquid rises in a capillary tube is inversely proportional to the radius of the tube; that is, the finer the tube, the greater the height to which the liquid rises. This law, which is known as Jurin's law, may be beautifully illustrated by placing two clean glass plates in a vessel of water, the plates being in contact along one edge, and separated a short distance on the opposite side, so that a wedge-shaped space is left between them. If these plates are then immersed in water, the water rises by capillarity between them, rising theoretically, at least, to an infinite height at the edge where the plates come in contact, and diminishing in height as the distance from the edge increases, so that the height to which the water rises is everywhere inversely proportional to the distance between the plates. The water thus forms a beautiful curve, known as the equilateral hyperbola, in the wedge-shaped space between the plates, the edge where the plates come together and the surface of the water in the containing vessel being the asymptotes.

It is a well known fact that when a soap bubble is suspended for a few moments from the pipe from which it was blown, the bubble contracts, and forces the air out through the pipe. This familiar phenomenon is due to the fact that the outer surface of the bubble is convex and therefore has a great normal molecular pressure inward, (p. 220), while the inner surface is concave and has a smaller normal molecular pressure outward. The pressure due to this surface film is therefore greater from the outside inward than from the inside outward, and the bubble accordingly contracts with a force equal to the resultant of these two pressures.

To a similar cause may be ascribed the movement of a drop of liquid when placed in a conical tube. If the liquid wets the tube, Fig. 149, the surface is concave at each end of the drop, the curvature being greater towards the smaller end. On the surface of the liquid towards the smaller end, the pressure due to this very concave film is less than that upon the other end, hence the resultant force due to the unequal pressure of the surface film upon the two ends of the drop is towards the smaller end of the tube. The drop, therefore, moves from a part of the tube where the cross-section is large towards the part where the cross-section is less. If, however, the liquid does not wet the tube, as in the case of mercury and glass, the surface of the drop is convex at both ends, Fig. 150, the curvature of the convex surface being greater towards the smaller end of the tube. The normal pressure due to the surface film, accordingly, is greater on the end of the drop having the sharper convexity, hence the resultant pressure is towards the larger end, and the drop moves towards the part of the tube where the cross-section is greater.

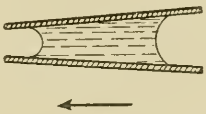
FIG. 149. Movement of water in a conical glass tube.

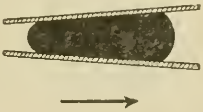
FIG. 150. Movement of mercury in a conical glass tube.

205. Attractions and Repulsions Due to Capillary Phenomena. Very interesting phenomena of attraction and repulsion are observed when various objects are floated upon this surface film of a liquid. Two objects that are wet by the liquid, as two pieces of wood floating upon water, will attract each other, or two objects that are not wet by the liquid, as two sewing needles likewise floating upon water, will also attract each other, while two objects, one of which is wet by the liquid and the other is not, will repel each other. In the same way, two glass plates suspended by threads and partly immersed in water will attract each other, or two glass plates partly immersed in mercury will also attract each other, while two plates, one of which is wet by the liquid while the other is not, will repel each other. The cause of these phenomena in all cases is purely hydrostatic.

In the first instance, when wetting occurs, Fig. 151, the liquid between the plates, being surmounted by a concave meniscus, stands above the general level, and the pressure within the elevated column must accordingly be less than that upon the exterior (p. 221). That is, if *C* is a point between the plates in the same plane as the general level, the pressure at *C* must be the same as that at the surface *AB*. Because of the concave meniscus, any point above *C*, as *D*, is at a pressure less than this amount by the weight of the column of liquid *CD*. Since the pressure between the plates is thus less than the exterior pressure, the resultant pressure is inward, and the plates are pushed together with a force which is equal to the weight of the column of liquid from *C* up to the curved surface.

In the second instance, where the convex meniscus between the plates stands lower than the general level, Fig. 152, the pressure at *C* is the same as that for points in the same horizontal plane *DE*. Since there is no liquid, and therefore no pressure, above *C*, the two plates are pushed together because of the exterior hydrostatic pressures *AD* and *BE*.

In the third instance, Fig. 153, where one plate is wet by the liquid while the other is not, the elevation of the liquid upon the wet plate lessens the depression upon the plate that is not wet, and this depression upon the latter plate lessens the elevation of the liquid upon the former. Accordingly, the liquid between the plates stands lower against the one that is wet, and higher against the one that is not wet, than it would if each plate were alone. But we have just seen that the greater pressure exists upon the side of the lower level in the case where the plate is wet by the liquid and upon the side of the higher level in the case where the plate is not wet. Therefore, since the pressure outward upon each plate is greater than the pressure inward, the plates move apart in obedience to the greater pressure in this direction.

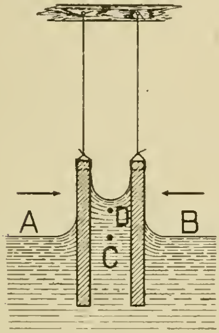

Fig. 151. Attraction between plates wet by a liquid.

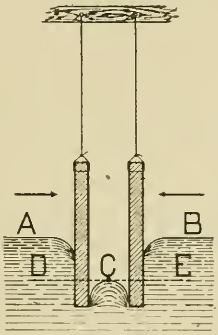

Fig. 152. Attraction between plates not wet by a liquid.

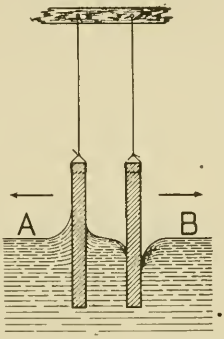

Fig. 153. Repulsion between a plate which is wet by a liquid and one which is not wet.

Another illustration of the attraction between bodies due to capillarity is found in the force with which two plates attract each other when they are pressed together with a drop of water or other wetting liquid between them, Fig. 154. The water is thus flattened out into a thin sheet, the edges of which are everywhere concave outward with a meniscus of very great curvature, due to the exceedingly small distance between the plates. But we have just seen (p. 220), that the pressure within a liquid bounded by a concave surface is

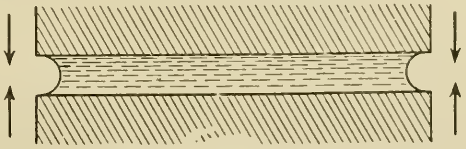

Fig. 154. Attraction between plates wet by a film of liquid.

less than the normal amount by the weight of a column of the liquid the height of which is directly as the curvature of the meniscus, that is, in this instance, inversely as the distance between the plates. Since the pressure upon the outside of the plates is one atmosphere, the pressure in the liquid between them is therefore less than this amount, and the plates are, accordingly, forced together with a pressure which is the resultant of these two.

If the plates are entirely submerged in the water, the meniscus disappears, the pressures within and without are then the same, and they no longer attract each other.

206. The Plateau Figures. Although totally blind, Plateau conceived the idea of studying the effect of surface tension by withdrawing wire frames from a soap solution, and these were then found to be covered with the most beautiful and delicate films, Fig. 155. These films, which consist almost entirely of the superficial layer of the liquid, give a most excellent opportunity for studying the surface tension without the necessity of considering the main body of the liquid. Figures of marvelous beauty and symmetry are obtained by the firms connecting the edges of these frames, the tendency of these films to contract reducing their area to the minimum surface. However complicated the figure may be, it is interesting to observe that, in

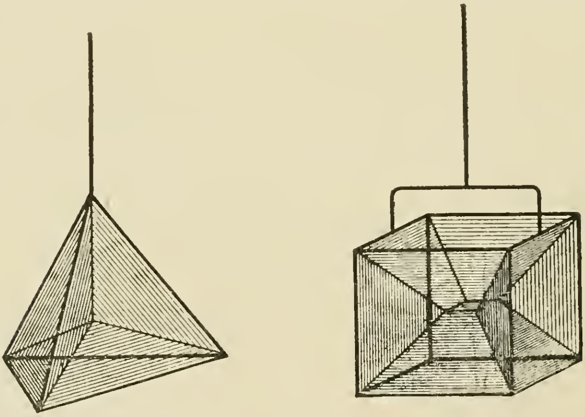

FIG. 155. Plateau's soap films.

each case where these films intersect, it is impossible for more than three films to meet in a line, or for more than four lines, that is, six films, to meet at a point, but in all cases the angles formed by the intersecting films and lines are absolutely equal. If a film breaks, so that for a moment this number is exceeded or the remaining angles are unequal, the condition is unstable, and a shifting occurs until the former relation again prevails.

CHAPTER XIV.

VORTICES AND DIFFUSION.

207. The Rotational Motion of Fluids—Vortices. A vortex consists of particles of a fluid in rotation immersed in a fluid that does not rotate.

The higher mathematical analysis leads to many remarkable facts connected with the rotational movement of fluids, the results of which only can be presented here.

1. A vortex in a perfect fluid always consists of the same fluid particles wherever in that fluid the vortex may move.

2. The strength of a vortex is the product of the cross-section of the vortex and its rotational, or angular, velocity, and is a constant quantity for that vortex. As the cross-section of the vortex diminishes, the rotation, accordingly, increases in the same proportion. A vortex, therefore, can never end within a finite fluid, but must either terminate at each end in the surface of the fluid, or be bent into the form of a ring, the two ends being brought together. Hence vortices are either linear or ring-shaped, and may be knotted or twisted in any manner whatever.

3. While the medium surrounding the vortex does not rotate, it nevertheless transmits the motion from one vortex to another. Therefore,

4. A single linear vortex remains stationary.

5. Two rectilinear vortices will revolve about their common center of mass.

6. A ring-shaped vortex moves forward along the axis of the ring in the direction in which the particles of the fluid on the inner surface of the ring are moving, Fig. 156.

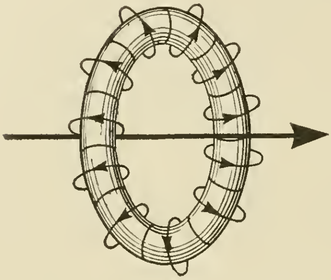

FIG. 156. A vortex ring.

7. The fluid enclosed by the ring, although not rotating, moves forward in the direction in which the ring is moving, and with a velocity greater than the ring.

8. If a vortex ring strikes an immovable wall, the diameter of the ring increases, the rotational velocity decreases, and the ring spreads out, as it were, against the wall. This is also true of each vortex when two vortices meet, moving along the same axis in opposite directions.

9. When two vortices of the same or nearly the same, strength move along the same axis in the same direction, one a little behind the other, the one in advance widens out and moves more slowly, and the one in the rear contracts and moves more rapidly and passes through the first. The second, now being in advance, widens out and moves more slowly, while the first, which is now in the rear, contracts and moves more rapidly, and so passes through the second. The one in the rear, therefore, always overtakes the one in advance and passes through it, and in a perfect fluid this play would be maintained indefinitely.

10. Since a vortex in a perfect fluid always consists of the same fluid particles, it follows that in such a fluid a vortex can never be destroyed, and for the same reason an act of creation is required to form a vortex. If a vortex in a perfect fluid exists, it has, therefore, always existed, and is indestructible.

208. The Vortex Atom. The results here set forth were discovered mathematically by Helmholtz in the year 1858. It remained, however, for Sir William Thomson, the late Lord Kelvin, to recognize the fact that a vortex resembles matter, in that it can neither be created nor destroyed, and upon this hypothesis was founded the theory of the vortex atom. This theory assumed that an atom of matter consists of vortex rings in the ether, and that one kind of matter differs from another because the vortex ring in one is knotted, or twisted, or looped with another vortex ring, like the links of a chain, while the vortex ring in another kind of matter is knotted or twisted differently, or is linked with a second vortex ring in some other way. There is, however, no experimental evidence in support of this theory of matter, and the hypothesis of the vortex atom has now given place to the theory that the atom must be electrical in its structure, possibly not unlike the one conceived by J. J. Thomson, to be described later.

209. Diffusion. If the lower part of a tall jar is filled with a heavy gas such as carbonic acid, and the upper part with a lighter gas such as hydrogen, it will soon be found that the two gases have not remained separated by their specific gravities, but have mingled completely, and that every part of the jar contains just as much of each gas as every other part.

If a solution of a heavy salt is poured through a tube down to the bottom of a vessel containing water, the greater density of the salt would always tend to keep it below, and the lighter water above. After a short time it will be found, however, that the particles of the heavy liquid have wandered into the upper part of the vessel, while the particles of the lighter water have descended and mingled with the heavier solution, and that ultimately the mixture will be uniform throughout. This tendency of gases thus to mingle, as well as the similar tendency of liquids likewise to mingle, is called diffusion, and is due to the fact that the particles of liquids as well as gases are in a state of continuous and irregular motion (p. 3).

210. Osmosis. The process of diffusion takes place in the manner described when nothing separates the two gases or liquids. If, however, a porous membrane or partition is placed between two gases, it will in general be found that the pores of the partition will permit one of the gases to pass more freely than the other, and the gases will still diffuse as before, though more slowly. If unglazed earthenware is used as such a porous partition, hydrogen, for instance, will pass through the pores more readily than air. In the same way the pores of a membrane, separating two liquids which would otherwise diffuse, will in general permit one liquid to pass more freely than the other. A membrane such as a piece of parchment paper, which permits the ready passage of water through its pores but prevents the passage of a dissolved substance, is called a semi-permeable membrane. When diffusion takes place through a membrane, the process is known as osmosis.

If a solution of sugar is placed in a vessel having a parchment bottom, and this vessel is placed in another containing pure water, the two liquids will tend to mingle through the membrane by osmosis. However, as the pores of the membrane separating them will allow molecules of water to pass but not molecules of sugar, water will continually pass from the outer vessel through the membrane into the sugar solution within. If the inner vessel is provided with a long vertical tube, the water, as it streams through the membrane, will cause the diluted sugar solution to rise in the tube, thereby exerting a hydrostatic counter-pressure which opposes the incoming water, and ultimately prevents its further passage.

211. The Osmotic Pressure. This hydrostatic pressure which is known as the osmotic pressure, and which is the pressure that finally prevents the futher flow of water through the membrane, may be taken as a measure of the ability of the two liquids to mingle. In many liquids the osmotic pressure is very great indeed, and may amount, as in the case of the sugar solution just mentioned, to many hundred atmospheres.

While the osmotic pressure differs greatly for solutions of different substances, the curious and interesting fact remains that the amount of this pressure depends only upon the number of separate particles of the dissolved substance present, and not at all upon their kind or nature. Two different solutions, therefore, will give rise to the same osmotic pressure, provided the number of separate particles of dissolved substance in each is the same. In all cases, moreover, the substance exerts an osmotic pressure in solution equal to the gaseous pressure which it would exert, if the same amount of substance were volatilized and so became a gas, and this gas were compressed into a volume equal to the volume of the solution.

HEAT.

CHAPTER XV.

EXPANSION.

212. Heat a Form of Energy. Until the closing years of the eighteenth century, heat was thought to be an invisible, weightless substance, called caloric, which was concealed in all bodies, and which, when set free, or released, produced a rise of temperature. Fallacies in this theory were soon discovered; for it was noticed that if this theory were true all bodies must contain practically an unlimited supply of heat, because when a small object, like a button, was rubbed upon a rough surface, an apparently inexhaustible quantity of heat could be developed.

It was not until the year 1798, however, that the true theory regarding the nature of heat was firmly and fully established. In that year, Count Rumford, in the city of Munich, as he was engaged in boring out brass cannon, observed that the metal of the cannon, the brass chips torn away, and also the boring-tool, became heated. Inasmuch as this process was carried on in a room of ordinary temperature, there was no source from which heat could have been derived other than something connected with the boring itself. Indeed, when the cannon was placed in an oak box filled with 18.7 lbs. of water at 60° F., and after the boring had continued for two and a half hours, the temperature of the water was raised to the boiling point. Count Rumford concluded that this enormous quantity of heat could have been produced only by the work done in tearing apart the particles of brass. Moreover, he then saw clearly that heat is developed when mechanical energy is expended, and, therefore, that heat must be a form of energy.

A few years later, Sir Humphrey Davy performed a similar experiment whereby heat was developed as a result of mechanical energy expended. He arranged a mechanism so that two pieces of ice could be rubbed together under the receiver of an air pump, the experiment being performed in a room at a temperature below the freezing point of water. As a result of the work thus done, the ice was partly melted; and as the melting of ice requires the application of a large quantity of external heat, it was seen that the heat thus produced could not in any way have been obtained from the ice itself. Hence it was proved that the energy expended in overcoming the friction of one piece of ice upon another was transformed into heat.

These two experiments illustrate beautifully the theory regarding the nature of heat which is held at the present time. This theory regards heat simply as one of the many forms of energy, and, indeed, it is the form that always appears when mechanical energy is expended in overcoming friction. It was not until the year 1842, however, that Mayer, who first announced the doctrine of the conservation of energy, foresaw that the energy consumed as heat must be considered when the attempt is made to show that the total energy in any closed system remains constant. Mayer found that water could be raised in temperature by shaking, and thus in a measure he anticipated the classic experiments of Joule, to be described later. Moreover, the action of the steam engine illustrates the fact that the energy represented by the heat of combustion of the coal may be transformed into mechanical work. Inasmuch, therefore, as mechanical energy may be transformed into heat, and heat again transformed into mechanical energy, it is proved beyond a doubt that heat is simply a form of energy.

213. Heat is the Kinetic Energy of the Molecules. This fact being established beyond question, it is interesting to investigate the nature of the energy which heat represents. Everything points to the fact that the molecules of a body are not at rest, but are in continual motion (p. 3), and hence these moving particles possess kinetic energy (p. 120). The motion of these particles must not be considered as regular and orderly, but as extremely irregular and heterogeneous; and as a result of this confusion the particles continually strike against one another and rebound, only to strike against still other particles. It is thought, also, that the parts of the molecules themselves may have a motion of vibration or rotation about their centers of gravity.

Not only are the molecules in this state of continuous and irregular motion, but it is now believed that the electrons which have been disengaged from their parent atoms (p. 11), and which are thought to move freely with their enormous velocities among the molecules of matter, play no unimportant part in these molecular bombardments. Though of exceedingly small mass, these electrons have velocities so great that, on the average· their kinetic energy is nevertheless the same as that of the molecules. The molecules and the electrons must then be regarded as being in *thermal equilibrium,* that is, they are continually colliding with one another, and, in general, at each encounter transmitting energy from one to the other, the one having the larger amount transferring a portion of its store to the one having less.

Unless a change of state takes place in the body, the addition of heat increases the kinetic energy both of the molecules and of the electrons, and the temperature of the body rises. Indeed, when energy of any form is added to the body whereby its temperatures is raised, the kinetic energy of the particles is increased. For this reason a blacksmith by hammering may heat a nail red-hot, the energy represented by the blows of his hammer being expended in causing the molecules and the electrons of the nail to vibrate faster. When friction is developed by two bodies rubbing over each other, the work done in breaking down the little irregularities between the two surfaces is transformed into an increased kinetic energy of the particles of the bodies, and the heat thus produced is a measure of the work done against friction. If a gas is compressed, the work done may develop so much heat that an easily inflammable substance, such as tinder, may be kindled, as in the fire-syringe. When the flint and steel are used to "strike fire", the work done by the flint in tearing away little particles of the steel is so great that these particles are heated to a temperature at which they take fire and burn in the air.

Certain tribes of Indians are known to kindle a fire by the energy expended in rubbing together two pieces of wood. In fact, the conversion of mechanical energy into heat by this means may easily be effected by any one with the aid of the simple device known as the "fire-drill," the experiment furnishing one of the most beautiful and striking illustrations of the modern theory of heat. By means of a bow and a leathern thong, a tapering stick is rapidly rotated in a slight pit, or depression, in a board, the upper end of the stick being held in place by a cup-shaped cavity in an iron block. The wood, which is intensely heated and ground to a powder by the friction, collects in a notch extending from the edge of the board to the center of the pit, and may soon be fanned into a glowing coal.

If the quantity of water is placed in a tube which may be rotated rapidly, and if pieces of cork are held against the sides of the tube to produce friction, the work expended in overcoming the friction will be transformed into heat, and this will raise the temperature of the water to the boiling point. If an electric current is sent through a wire, the electrons in the metal are acted upon by a directing force which compels a majority of their number to move in a single direction. The energy which is thus expended in producing this directed motion reappears in the increased energy of the electrons, and therefore in the increased energy of the molecules when the electrons collide with them. A portion of the energy of the current is thus transformed into increased kinetic energy of the electrons and the molecules, and the temperature of the wire is raised.

These and many other experiments illustrate the great truth that heat is produced as a result of work performed, from which the inevitable conclusion is drawn that heat must be one of the forms in which energy appears. It is therefore not only a well-established fact that heat is a form of energy, but it is now believed that the heat of a body is the kinetic energy of its ultimate particles, that is, of its electrons and its molecules, and that the temperature of the body is simply a measure of this kinetic energy.

Whatever the effects of heat upon a body may be, they are simply such as would follow from the addition of energy to the body. Among these effects may be mentioned the elevation of temperature, expansion, change of state, chemical action, the production of an electric current, etc. Each of these effects will be studied in detail later.

214. Temperature. As we have seen above, the temperature of a body is the measure of the kinetic energy of the molecules and electrons. Accordingly, if two bodies are brought together, the temperature of one being higher than that of the other, the molecules and electrons of the one at higher temperature will communicate a portion of their motion, and therefore a portion of their energy, to the molecules and electrons of the other, and the temperature of the first will fall, while that of the second will rise. Kinetic energy has thus been transferred from the first body to the second, and heat has therefore flowed in the same direction between these two bodies. Temperature, accordingly, may be defined as the condition of a body that determines a flow of heat; and the difference of temperature between two bodies is the difference in the kinetic energies of their ultimate particles, due to which heat will flow from the body at higher temperature toward the one at lower temperature.

215. The Thermometer. Because it is the tendency of nearly all bodies to expand when heated, the amount of expansion may be taken as a measure of the temperature to which a body is raised. Any device thus for measuring temperature is known as a thermometer. The expansion of a liquid is in general much greater than that of a solid for the same rise of temperature, hence a liquid is commonly used as the thermometric substance. Owing to the fact that mercury has a very uniform expansion, a comparatively high boiling point, and a low freezing point, this is the material that is ordinarily employed in the manufacture of thermometers.

In the construction of thermometers a glass bulb, provided with a long, fine, uniform tube, is filled with mercury, the whole is then heated until the mercury overflows at the top, and in this condition the tube is sealed by melting the glass. As the mercury contracts upon cooling, a nearly perfect vacuum is left above the remaining column in the tube, and the rising and falling of this column, as the mercury within the bulb expands and contracts, may now be read from graduations placed upon the stem.

After the process of filling is complete, an interval of time, amounting often to years, must elapse before the graduations may be made, in order that certain molecular changes which occur in the freshly-blown bulb may be eliminated. Even after cooling, the glass resembles an extremely viscous fluid, and the molecules do not at once assume their final positions with reference to one another, so that for many years a slow contraction in the volume of the bulb takes place. Accordingly, the mercury gradually rises in the stem, as this contraction continues; and it is not until the thermometer has been thoroughly aged, by lying until this molecular change in the bulb practically ceases, that the instrument is ready for graduation.

216. The Fixed Points on the Thermometer. For a given change of temperature, the actual rise or fall of the mercury in the stem depends upon the volume of the bulb and the size of the capillary tube. As these dimensions are different in different instruments, and *in order to have an observed movement of the mercury column correspond to an exact interval of temperature,* it is necessary to determine individually for each thermometer the so-called *fixed points,* that is, the heights to which the mercury rises when the instrument is brought to two perfectly definite and constant temperatures.

It chances fortunately, that there are two temperatures, not only easily accessible, but also constant, under normal conditions, in all places and at all times. These temperatures are those at which ice melts in the air, and at which water boils under a barometric pressure of 76 centimeters. With these constant temperatures thus available, the fixed points on the barometer may now be located. The instrument is first imbedded in melting ice, and the point to which the mercury column falls is marked upon the stem. It is then placed in a bath of steam produced from water boiling under normal pressure, and the point to which the mercury thread then rises is likewise marked upon the stem. The distance between these two points corresponds to a perfectly definite range of temperature, and this interval may now be divided into any number of equal parts, called degrees.

217. The Thermometric Scales. There are in use two main systems of thermometer graduation, the Fahrenheit, F., and the Celsius, or centigrade C. On the Fahrenheit scale, Fig. 157, the melting point of ice is marked 32, the boiling point of water 212, and the interval between these two temperatures is divided into 180 degrees. The zero of this scale is thus 32 degrees below the melting point of ice. On the Celsius, or centigrade, scale, the melting point of ice is marked 0 and the boiling point of water 100, the interval between these points being divided into 100 equal parts, or degrees. Since the centigrade system of graduation is strictly decimal, it is by far the more convenient, and is used in all scientific work.

It is to be observed that 180° F. equal 100° C., or 9° F. equal 5° C., and that the zero of the Fahrenheit scale is 32 Fahrenheit degrees below the melting point of ice. The following relations therefore serve to transform temperatures from one scale to the other:

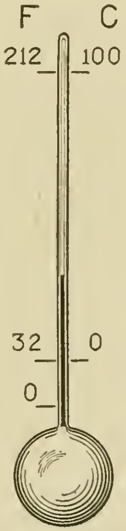

Fig. 157. The thermometric scales.

$$F^\circ = \frac{9}{5}\,C^\circ + 32.$$

$$C.^\circ = \frac{5}{9}\,(F.^\circ - 32.)$$

218. The Expansion of the Bulb. When heat is suddenly applied to a thermometer, the bulb expands almost exactly as if it were a solid ball of glass, and this occurs before the heat has had time to pass to the mercury within. Since the bulb now occupies a larger volume, while the volume of the mercury remains for a short time unchanged, the mercury must fall in the stem in order that the bulb may remain filled. The first effect, therefore, of suddenly applying heat to a thermometer is to cause the mercury to fall; and it is not until the heat has had time to pass through the glass and into the mercury that the latter begins to expand. The mercury now expands more rapidly than the glass and so rises in the tube, the apparent expansion thus indicated being the difference between the real expansion of the mercury and the expansion of the glass.

219. The Air Thermometer. The attempt has been made to use a gas instead of a liquid as the thermometric substance, because the expansion of a gas is much greater than that of a liquid for the same rise of temperature, and also because this expansion is perfectly uniform. But as the volume of a gas varies with every change of pressure, an air thermometer cannot be used without inconvenient and difficult barometric corrections. When these corrections are made, however, a gas proves to be the most reliable substance that can be used.

220. The Absolute Zero and the Absolute Scale of Temperature. When a quantity of gas is heated or cooled while the pressure to which it is subjected remains constant, experiment shows that the gas will expand $\frac{1}{273}$ of its volume at 0° C. for each degree C. it is heated, and will contract $\frac{1}{273}$ of its volume at 0° C. for each degree C. it is cooled. If, therefore, a mass of gas were cooled 273° below 0° C., the gas would contract to zero volume, provided the same law held true to this point. This we may assume to be the case with a perfect gas, although any known gas would change its state before this temperature were reached. The more rarefied a gas may be, however, the more nearly does it approach the behavior of a perfect gas in this respect.

An interpretation of this point may be obtained by recalling the fact that temperature is a measure of the kinetic energy of the molecules (p. 232). If a mass of gas at 0° C. is cooled

1° C., there is subtracted from it $\frac{1}{273}$ of the kinetic energy of the molecules of the gas at 0° C. Indeed, for each degree it is cooled there is withdrawn $\frac{1}{273}$ of the kinetic energy of the gas at 0° C. Hence, if the mass of gas were cooled 273 degrees, there would be withdrawn $\frac{273}{273}$ of the kinetic energy of the molecules of the gas at 0° C., and at this point the molecules would have no energy and the gas no volume. Therefore, the point —273° C., is the point of absolute rest of the molecules, and is called *the absolute zero. Temperatures on the centigrade scale, when measured from this point as zero, are called temperatures on the absolute scale, or absolute temperatures.* If t denotes ordinary temperatures on the centigrade scale, and T denotes absolute temperatures, then

$$T = t + 273.$$

221. The Law of Charles. The truth contained in the last paragraph, that is, the law expressing the fact that at constant pressure a mass of gas expands or contracts $\frac{1}{273}$ of its volume at 0° C. for each degree C. it is heated or cooled, is known as the law of Charles, or, sometimes, the law of Gay-Lussac. The law of Charles is commonly stated in the following words: *The pressure remaining constant, the volume of a mass of gas is directly proportional to its absolute temperature.*

If, however, a mass of gas is enclosed within rigid walls so that it can neither expand nor contract when heat is added or withdrawn, its volume remains constant throughout, but its pressure will vary as the temperature is changed. This may readily be seen by investigating the nature of the pressure of a gas.

From the modern theory of heat, which assumes that the molecules of a gas are in a state of rapid but irregular vibration, it is obvious that these particles must not only encounter frequent collisions with one another, but must likewise impinge upon the walls of the containing vessel, thereby striking a blow. The walls of a vessel in which a quantity of gas is enclosed are therefore considered as undergoing a continuous bombardment from these gaseous particles as they strike against it; and the pressure of an enclosed gas upon the walls of the containing vessel must be regarded as produced by the added impacts, or blows, delivered upon the walls by these particles. Accordingly, when the volume remains constant, the pressure of a gas must be directly proportional to the energy of its vibrating particles, and therefore directly proportional to the absolute temperature. From this point of view, a second form of the law of Charles may be expressed as follows: *The volume remaining constant, the pressure of a gas is directly proportional to its absolute temperature.*

From the above reasoning we see that if the volume V of a mass of gas changes to the volume V' while the absolute temperature T changes to the absolute temperature T', the pressure remaining constant, the first statement of Charles's law may be expressed more simply by the proportion,

$$V : V' = T : T' \text{ (the pressure remaining constant).}$$

But if the pressure P of a mass of gas changes to the pressure P' while the absolute temperature T changes to the absolute temperature T', the volume remaining constant, the second form of Charles's law may be written

$$P : P' = T : T' \text{ (the volume remaining constant).}$$

222. Combination of Boyle's Law and Charles's Law. By Boyle's law we have seen (p. 194) that, the temperature remaining constant, the volume of a mass of gas is inversely proportional to the pressure, and that it may be expressed by the relation

$$V : V' = \frac{1}{P} : \frac{1}{P'},$$

or

$$\frac{V}{V'} = \frac{P'}{P} \text{ (}T\text{ remaining constant).}$$

By Charles's law we have learned that, the pressure remaining constant, the volume of a mass of gas is directly proportional to the absolute temperature, and that this may be written

$$V : V' = T : T'$$

or

$$\frac{V}{V'} = \frac{T}{T'} \text{ (}P\text{ remaining constant).}$$

If, now the volume V of a mass of gas changes to a new volume V', while at the same time the pressure P and the absolute temperature T change to new values P' and T', respectively, the final relation between volume, pressure, and absolute temperature of the gas may be determined by conceiving the change of volume due to the change of pressure, and also the change of volume due to the change of absolute temperature, to take place *independently and successively.* That is, Fig. 158, the pressure P may be changed to its final value P' while the volume V changes by Boyle's law to some intermediate volume, V_x, the temperature remaining constant, so that

$$V : V_x = \frac{1}{P} : \frac{1}{P'} \text{ (Boyle's law, }T\text{ remaining constant).} \quad (1)$$

The absolute temperature is then changed from its original value T to its final value T', which causes the volume to change by Charles's law from its intermediate value V_x to its final value V', the pressure P' remaining constant, so that

$$V_x : V' = T : T' \text{ (Charles's law, }P'\text{ remaining constant).} \quad (2)$$

From (1),

$$\frac{V}{V_x} = \frac{P'}{P} \text{ (Boyle's law),} \tag{3}$$

and from (2),

$$\frac{V_x}{V'} = \frac{T}{T'} \text{ (Charles's law).} \tag{4}$$

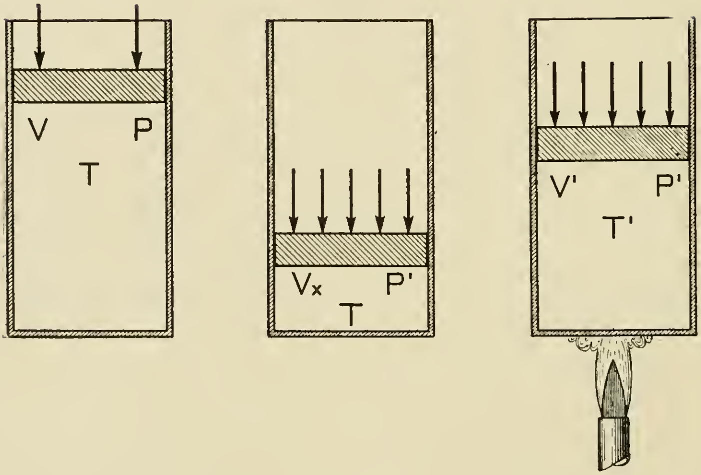

FIG. 158. Boyle's law and Charles's law operating independently and successively.

Multiplying (3) and (4) member for member, we have

$$\frac{V}{V'} = \frac{P'\,T}{P\,T'}, \quad \frac{1}{TT'}, \quad \frac{VPT'}{TT'} = \frac{V'P'T}{TT'}$$

from which

$$\frac{P\,V}{T} = \frac{P'\,V'}{T'}\,.$$

That is, for a given mass of gas the value of the expression

$$\frac{P\,V}{T}$$

does not change in consequence of this series of operations, and therefore remains constant. Letting R represent this constant, we have

$$\frac{PV}{T} = R,$$

or

$$PV = RT,$$

from which it follows that the product of the volume and the pressure of a mass of gas is always proportional to its absolute temperature

223. Expansion by Heat. The temperature of a body has been defined, (p. 234), as the measure of the kinetic energy of its molecules and electrons. When, therefore, its temperature is raised, the kinetic energy of its particles is increased, and the molecular bombardments and collisions take place more frequently and more vigorously than at a lower temperature. One result of this increased molecular and electronic energy is the crowding apart of the molecules; and therefore when its temperature is raised the body as a whole expands. Indeed, the effect of expansion is one so commonly following the addition of heat to a body that we may formulate the general rule, to which, however, there are numerous exceptions, that a rise of temperature causes the body to expand, and a fall of temperature causes it to contract. More briefly, this may be stated in the maxim that heat expands and cold contracts.

224. The Force Developed by Expansion. The force, moreover, developed by bodies in expanding and contracting is so almost irresistible that an allowance must be made for this effect in the erection of large buildings and in the construction of long steel bridges. For the same reason spaces are left between the ends of the rails on a railroad, so that when expansion occurs the rails may not be warped out of their true course. Frequently the walls of a building are straightened by passing a long iron rod from one side to the other, then heating this rod and screwing it up tightly by nuts on the outside. The force of contraction manifested by the rod as it cools, is sufficient to draw the walls back again into their proper position. An idea of this enormous molecular force may be obtained by screwing a heated iron bar in a heavy metal frame, one end of the bar being held in place by a cast iron rod. As the bar cools, contraction occurs, and the force with which this takes place, being far greater than the strength of the iron rod, causes the latter to break.

225. Differences in Expansion. Bodies differ greatly in the amounts by which they expand or contract for a given change of temperature, this variation in length being greater, for instance, with zinc than with steel. Accordingly, if a strip of zinc and one of steel are riveted together and heated, the zinc, in expanding more than the steel, will cause the whole to bend in the arc of a circle, the zinc being upon the outside. If the compound strip is cooled, the zinc will contract more than the steel, and the strip will bend in an arc of opposite curvature with the zinc upon the inside. If, now, one end is held fast, the other will move through a considerable range, depending upon the degree of the heating or cooling. At any desired point of its path, it may be made to close or open an electric circuit controlling a radiator, so that the heat of a room may thus be automatically regulated, the compound strip then forming what is known as a thermostat. Upon this principle is also constructed the metallic thermometer. Here strips of two different metals are riveted or soldered together, the whole then coiled into a spiral form with one end rigidly held. A pointer attached to the other end may be made to move over a scale which has been graduated by comparison with a mercurial themometer, and temperatures may be thus indicated.

226. Molecular Forces may Mask Expansion. In many of the instances where an apparent exception to the general rule of expansion by heat occurs, it is undoubtedly true that other molecular forces, as, for instance, the forces of elasticity and crystallization, are greatly changed by a rise of temperature, and the changes that take place in these forces due to the rise of temperature may to some extent mask the effect of expansion. It is probably on this account that stretched rubber and certain crystalline bodies actually contract as the temperature is raised.

227. Coefficient of Linear Expansion. When a body expands with a rise of temperature, and we have to consider the resulting increase in volume, we have to deal with what is known as the voluminal, or cubical, expansion of the body. If, however, the body is in the form of a long, thin bar, or rod, and this is heated, the expansion of the bar in the direction of its length is so much greater than that in the direction of its breadth or its thickness that the latter effects may be neglected altogether, and we have only to take account of its linear expansion.

The amount of the linear expansion of a bar is found to be proportional to its length, to the rise in temperature, and to a certain constant, called the coefficient of linear expansion, which depends upon the nature of the material of which the bar is made. If e denotes the actual increase in length of a bar when its temperature is raised, l_0 the length of this bar at 0° C., t the temperature on the centigrade scale to which the bar is heated, and α the coefficient of linear expansion, it follows that

$$e = l_0 \, \alpha \, t.$$

If the bar is exactly 1 centimeter long, ($l_0 = 1$), and if this bar is raised in temperature 1° C., ($t = 1$), then the coefficient of linear expansion equals the actual expansion which takes place in the length of the bar. That is, under these conditions,

$$e = \alpha.$$

The coefficient of linear expansion of a substance, therefore, is the increase in the length of a bar of the substance 1 centimeter long when its temperature is raised from 0° C. to 1° C. Accordingly, the length of a bar at any temperature may be calculated if its length at 0° C., the temperature to which it is raised, and its coefficient of linear expansion are known; for the length l_t of a bar at any temperature t, is equal to its length at 0° C., plus the amount it has expanded, Fig. 159, or

$$l_t = l_o + e,$$

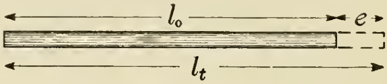

FIG. 159. Linear expansion.

from which

$$l_t = l_o + l_o \alpha t,$$

or

$$l_t = l_o (1 + \alpha t).$$

The quantity within the parenthesis is known as the factor of expansion. The length of a bar at any temperature may therefore be determined if its length at 0° C. is multiplied by the factor of expansion.

228. The Coefficient of Cubical Expansion. The coefficient of cubical expansion of a substance is the increase in the volume of 1 cubic centimeter of that substance when heated from 0° C. to 1° C. Accordingly, if V_o is the volume of the substance at 0° C., and V_t is its volume when heated to $t°$, β being its coefficient of cubical expansion,

$$V_t = V_0 + V_0 \beta t,$$

or

$$V_t = V_o (1 + \beta t). \qquad (1)$$

An interesting relation exists between the coefficient of linear expansion α, and that of cubical expansion β, and may be determined in the following manner: Since the volume of a cube is the length of one side of that cube raised to the third power,

$$V_o = l_o^3. \qquad (2)$$

At any other temperature t, the volume V_t may be expressed by

$$V_t = l_t^3 = l_0^3\ (1 + \alpha\, t)^3,$$

where α is the coefficient of linear expansion. Developing the quantity in the parenthesis,

$$V_t = l_0^3\ (1 + 3\,\alpha\, t + 3\,\alpha^2\, t^2 + \alpha^3\, t^3).$$

But as α is an exceedingly small quantity, the terms containing α^2 and α^3 may be neglected, and therefore

$$V_t = 1_0^3\ (1 + 3\,\alpha\, t).$$

Substituting the value of l_0^3 from equation (2),

$$V_t = V_0\ (1 + 3\,\alpha\, t). \qquad (3)$$

By comparing equations (1) and (3), we see that

$$\beta = 3\,\alpha.$$

Therefore the coefficient of cubical expansion is very nearly 3 times the coefficient of linear expansion.

229. The Anomalous Expansion of Water. Water has a most remarkable and abnormal expansion, the effects of which are so far-reaching that we are undoubtedly indebted to this peculiarity, in part, at least, for the mildness of the climate in the temperate zone. A mass of water at 0° C., when heated, contracts in volume until the temperature of 4° C. is reached, after which, as the temperature rises further, it begins to expand, and continues to expand until the boiling point is reached. At the temperature of 8° C., the water has attained practically the same volume it had at 0° C. This is shown in Fig. 160, where the vertical distances represent the changes in the volume of a mass of water, and the horizontal distances indicate the temperatures. At 4° C. a mass of water is thus seen to have its smallest volume, and therefore its greatest density (p. 77).

As a mass of water in a river or lake is cooled from ordinary temperatures down to 0° C., it contracts and becomes denser until the temperature of 4° C. is reached, at which point it attains its maximum density. This denser, colder water, therefore, sinks to the bottom, and warmer, lighter water takes its place at the surface. This in its turn, loses some of its heat by radiation into the air, and so cools and sinks, and this process continues until the whole mass of water in the entire lake is at 4° C. Then as the surface layers lose still more heat by contact with the cold air, they expand and become lighter, and so float upon the warmer, denser water beneath. The water at the surface continues to lose heat and become lighter, while the lower layers and the water at the bottom remain at 4° C., until the surface falls to 0° C., when it freezes. The warmer, denser water which remains at the lower levels, is able to cool, and therefore freeze, only as heat is carried away from it by conduction through the layer of ice already formed. This, however, is a slow process; hence the ice cannot increase rapidly in thickness, and a comparatively thin layer forms at the surface.

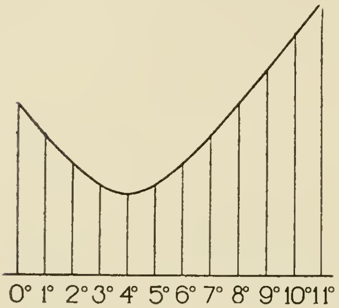

FIG. 160. The maximum density of water occurs at 4° C.

CHAPTER XVI.

THE TRANSFERENCE OF HEAT.

230. Methods of Heat Transference. It is a matter of common observation that heat may be communicated from one body to another of lower temperature, or transferred from one part of space to another. The methods by which this transference takes place are three in number, and are known as conduction, convection, and radiation.

231. Conduction. Conduction is the traveling of heat from particle to particle, and its transference in this way from one portion of a heated body to another. When one end of a bar of metal is heated, the molecules at this point, and, it is believed also, the electrons as well, vibrate with increased energy, and the temperature of the heated portion rises. As these particles have their kinetic energy increased, they set into more rapid vibration the molecules and the electrons adjacent to them, these, in turn, impart additional energy to the particles beyond them, and so on, until even distant portions of the bar attain a higher temperature. Heat has thus travelled along the bar, being communicated from particle to particle in the form of increased kinetic energy of the particles throughout the length of the bar.

Substances differ greatly in their power thus of conducting heat, the metals being better conductors than other substances. This may be illustrated by wrapping a piece of writing paper tightly around a compound bar, one half of which is wood, while the other half is brass or copper. If the whole is now plunged into the flame of a bunsen burner, the paper will be charred around the wood but uninjured around the metal, the exact line where the two are joined being in this way plainly marked. The copper conducts away the heat so rapidly that the paper is kept cool, while the wood is so poor a conductor that the heat remains in the paper, and scorching occurs.

Even among the metals there are wide differences in conductivity, silver and copper being the best. It is because of the rapid conduction of heat by metals that a piece of wire gauze may be lowered over a flame and the latter fail to burn above, the metal of the gauze conducting away the heat so rapidly that there is left insufficient heat to ignite the gas. For the same reason a piece of gauze may be lowered over a jet of gas, and the gas lighted above. The flame above the gauze does not ignite the gas below, because the heat from the flame above has been conducted away by the metal of the gauze.

Sir Humphrey Davy made use of this principle in the construction of the Davy safety lamp for use in mines where inflammable gases are prevalent. Here the flame of the lamp is surrounded by fine wire gauze, through which the explosive gases of the mine are free to pass. Little explosions of the inflammable gases indeed take place within the lamp; but the flame within cannot pass through the meshes of the gauze and so ignite the explosive gases without, because the metal conducts away the heat so rapidly and completely that the gases on the outside are not heated to the point of ignition. In this way the miner may carry this lamp into parts of the mine filled with explosive gases, where it would be dangerous or impossible to go without this device.

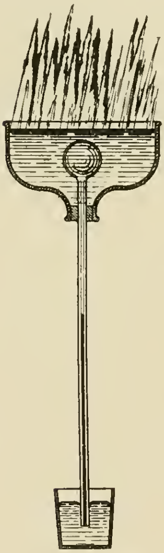
Fig. 161. Water is a poor conductor of heat.

Liquids have a very low power of conductivity for heat, as may be illustrated by igniting a layer of turpentine or ether on the surface of water, beneath which is placed a delicate air thermometer, Fig. 161. Although the bulb of the thermometer is but a few millimeters below the surface of the water, and therefore removed but a short distance from the intense source of heat above, the index of the thermometer hardly moves, so small is the conductivity of water for heat. For the same reason water may be boiled in the upper part of a test tube, while a piece of ice, attached to a weight to keep it submerged, remains unmelted at the bottom.

Gases differ greatly among themselves in their power of conducting heat; and, as might be anticipated, those gases, the particles of which are in the most rapid motion, conduct heat the most readily. Thus hydrogen, whose molecules have a higher velocity than those of any other gas, conducts heat far more readily than does carbonic acid gas, while a perfect vacuum does not conduct heat at all.

It is to the thermal insulation produced by a vacuum that is due the efficacy of the Dewar flask for holding liquid air. This exceedingly cold substance is contained in a thin glass flask which is surrounded by an outer flask, Fig. 162, the space between the two being as nearly as possible a perfect vacuum. Conduction of heat through this highly exhausted space is thus reduced to a minimum; and if the further precaution is taken to deposit a thin coating of silver upon the vessels, the energy, which would otherwise be transmitted by radiation (p. 253) through the vacuum to the inner vessel, is reflected away, and the liquid air, thus protected from almost all access of heat, may be preserved many days. It is upon this principle that is constructed the "thermos bottle", which has recently come into general use, and which is effective in keeping liquids hot or cold for long periods of time.

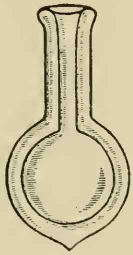

FIG. 162. The Dewar flask.

232. Convection. Convection is the transference of heat by the actual motion of the heated body, and can occur only in liquids and gases. This process, which is a purely mechanical one, is usually effected by currents, or circulation, produced by the differences in specific gravity between the heated and cooled portions of the body. Thus, in a room, the air in contact with the stove or radiator is heated, and, as it expands, it becomes less dense than the surrounding air, and so rises, while the cooler air from the sides flows in to take its place. Currents of warm, light air, therefore, continually rise from the source of heat and move to distant parts of the room, distributing heat to portions far removed from the source of heat.

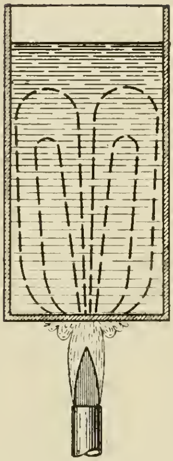

FIG. 163. Convection currents.

The circulation of water may be studied in the same way by scattering a few grains of sawdust or a little coloring matter in a beaker of water heated by a flame. Currents of heated water will thus be seen to ascend from a point immediated over the flame, Fig. 163, while currents of cooler, denser water will descend at the sides and move towards the source of heat. In a similar way, a small coil of platinum wire at the bottom of a tank of water may be heated by an electric current and so warm the water in its immediate neighborhood. The heated water expands, thus becoming less dense than the water around it, and so rises, while cooler water from the sides and bottom of the tank flows towards the coil, and this in turn expands and rises in the same way.

233. Radiation. Radiation is the term used to denote a transference of energy by means of a vibratory motion of the ether. This phenomenon is believed to have its origin in the electrons of the heated body, which are thought to be vibrating or rotating many thousands of million times in a second. The vibratory or rotary motion of the electrons disturbs the adjacent particles of the ether, these particles disturb other neighboring particles, and so on, so that a wave, which thus starts at the heated body, is propagated in all directions through the ether with the velocity of light.

This energy of vibration of the ether is not heat, and does not become heat until these waves encounter a substance through which they cannot pass freely. When they are stopped by an opaque body, their motion of vibration ceases, and the energy represented by this motion is expended in setting in vibration the molecules of the opaque body itself, and so the latter becomes warm. It is not, therefore, until the waves of the ether are retarded, and their energy transformed into the energy of vibration of material particles, that this energy becomes apparent as heat.

It is in this way that vibratory, or radiant, energy comes to us from the sun. The vibrations originating in the sun are propagated in all directions through the ether, and the entire amount of ether throughout space is thrown into vibration. This energy of vibration does not become the energy of heat until the vibratory motion is arrested by an opaque body, such as the earth; hence the regions of inter-stellar space, although transmitting an unlimited amount of energy, are intensely cold.

The experiment due to Professor Tyndall of concentrating the sun's rays by a lens made of ice beautifully illustrates the principle that radiant, or vibratory, energy passes through the medium that transmits it without heating that medium. If a block of ice is formed in the shape of a lens, this may be used for bringing to a focus the rays of the sun and igniting a substance like phosphorus or tinder. The vibrations of the ether, finding but little in the ice to obstruct their motion, pass through the ice with but slightly diminished energy, and it is not until these vibrations are stopped that their energy is transformed into heat. In this way a lens of ice may be used as a "burning glass".

234. The Radiometer. The little instrument known as the radiometer depends for its action upon the principle that heat is developed when the vibratory motion of the ether is arrested. This consists of four little vanes of mica or aluminium, mounted so as to turn freely on a pivot, Fig. 164. one side of each vane being blackened, while the other is left bright. The whole is then mounted in a glass globe from which nearly all the air has been exhausted.

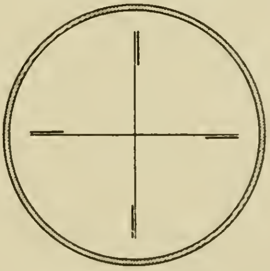

FIG. 164. The radiometer.

When no radiation falls upon the vanes, that is, when the instrument is kept in the dark, the bright and the black faces are at the same temperature. When, however, radiant energy from the sun or other luminous source falls upon them, the blackened vanes check the vibration of the ether and convert their energy into heat, thereby raising the temperature of these faces, while the bright vanes reflect away this energy, and thus remain at a constant temperature. Thus, the only effect of radiation is to raise the temperature of the blackened surfaces, and to maintain unchanged the temperature of those that are bright.

The rotation of the vanes, however, is occasioned by the unequal reaction of the air particles, as they rebound from the blackened and the bright surfaces after colliding with them. A particle of air in the partially exhausted globe will bombard the bright surface of a vane and rebound with the same velocity as that with which it approached, since it receives no energy from the vane. On the other hand, a similar air particle will rebound from a blackened surface with a velocity greater than that with which it approached, since it has been in contact for an instant with this heated surface, and has received energy from it. The impact of an air particle upon a bright surface is counterbalanced by a similar and equal impact upon a blackened surface; but the reaction upon the blackened surface, as a particle rebounds, is greater than that upon the bright surface, because the particle leaves the blackened surface with an increased velocity, and therefore pushes back with an additional force as it leaves. The blackened faces of the vanes thus recede and a rapid rotation is produced, the momentum of the vanes and the increased momentum of the air particles being equal in amount and opposite in direction.

235. Reflection, Transmission, and Absorption. When radiant energy falls upon a body, this energy is, in general, divided into three parts, depending upon the nature of the body. One portion is transmitted, that is, passes on through the body, another portion is reflected, and a third portion is absorbed.

If the body is polished, and if it offers but little obstruction to the passage of the ether waves, the radiant energy is largely transmitted through the body, and the body is said to be transparent to these waves, or diathermous. Again, if the radiant energy does not penetrate the opaque body, but is turned back at the polished surface, the energy is reflected, and the body is a good reflector. If, however, the nature of the body is such that only a small portion of the incident energy is reflected or transmitted, and therefore a large part is absorbed, the body is said to be a good absorber. Rock salt is one of the most diathermous of all known substances, the polished surfaces of metals are good reflectors, silver being the best, while the substance that most completely absorbs all radiations falling upon it, and therefore the one that proves to be the best absorber, is lampblack.

236. Radiation—Prevost's Theory of Exchanges. If the radiations from a source of energy, such as a gas flame, fall upon a delicate thermometer or thermopile, a portion of this radiant energy is absorbed, and the temperature of the thermometer rises. If now a piece of ice instead of the gas flame is held near, the mercury will be observed to fall, showing a lower temperature, and apparently indicating that cold has been radiated from the ice. When we remember, however, that radiation is the process of transmitting energy by the vibratory motion of the ether, we see at once that it is only energy, and not a lack of energy, which can be transmitted in this way, and therefore that it is impossible for the ice to radiate cold. Although this experiment is most simple, it is nevertheless most instructive, because it gives us an insight into the conditions of bodies at ordinary temperatures as regards radiation.

From the latter experiment it is clear that the thermometer lost energy due to the presence of the ice, since its temperature was lowered, and it is also clear that this loss of energy could not have been radiated from the ice. The conclusion is therefore evident that the thermometer lost energy by the radiations which it itself sent out to the ice, and that this loss was greater than it would have been if a body of the same temperature as the thermometer had been used instead of the ice.

Experiments of a similar nature have led to the belief that all bodies are continually radiating energy to other bodies, and also continually receiving energy by radiation from other bodies. If a body is of higher temperature than surrounding objects, it sends out by radiation more energy than it receives in the same way from these objects, and therefore it grows cold, while the surrounding objects grow warm. Similarly if a body is colder than surrounding objects, it receives more energy by radiation from these objects than it sends out to them, hence the body grows warm, owing to the excess of energy which it receives by radiation over that which it sends out in the same way. Furthermore, if a body is at the same temperature as surrounding objects, we must not conclude that it does not radiate or receive radiations, but must rather believe with Prevost that it is continually radiating energy to the walls of the room and to objects in its neighborhood, and that it is also constantly receiving energy by radiation from these objects. Moreover, the amount which it thus loses by radiation is exactly equal to the amount it receives in the same way, so that its temperature does not change.

The experiment with the ice is therefore at once interpreted, and we now see how it is indeed true that the ice does not radiate cold, but that it radiates energy, and that the thermometer also radiates energy. Inasmuch, however, as the thermometer is at a higher temperature than the ice, it radiates a greater amount of energy to the ice than it receives by radiation from the ice, hence its temperature falls. Radiation must, therefore, be considered as taking place at all temperatures from all bodies. If the bodies are at different temperatures, the effect of radiation will be to bring them to the same temperature; if the bodies are initially at the same temperature, a constant interchange of energy is taking place between them, and the bodies remain at the same temperature due only to the fact that each body continually receives as much energy by radiation from adjacent bodies as it itself sends out by radiation to these bodies.

237. Equality of Radiating Power and Absorbing Power. Possibly a clearer view of this exchange of radiations may be obtained if we think of a body suspended in the center of a hollow sphere, Fig. 165. If the inner body is at a higher temperature than the surrounding sphere, it will radiate more energy to the sphere than it receives in return, and the two will be brought to an equality of temperature. If the inner body is colder than the sphere, it will receive more energy than it radiates, and again the two will come to a common temperature. If the inner body is at the same temperature as the sphere it must radiate as much energy as it receives, otherwise it would continually grow warmer; moreover, it must also receive as much energy as it radiates, or it would continually grow colder. The ability of a body to radiate, therefore, is exactly equal to the capacity of the same body to absorb radations, that is, the radiating power of a body is equal to its absorbing power, where the radiating power of a body is defined as the amount of energy radiated per second per square centimeter, and the absorbing power of a body is defined as the amount of energy absorbed per second per square centimeter. We understand, therefore, how it is that lamp-black, which is the best absorber, is also the best radiator, while a body which is a poor absorber or a good reflector, is also a poor radiator.

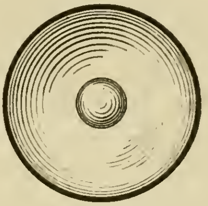

FIG. 165. Radiation and absorption are equal.

238. Leslie Cube. The great difference that exists between various surfaces as regards their radiating power, and therefore, also, as regards their absorbing power, may be beautifully illustrated by the device known as the Leslie cube. This is a hollow metallic cubical box, polished upon one of its lateral surfaces, left in a roughened state upon another, covered with lamp-black upon a third, and coated with a paint or material to be investigated upon the fourth. The four faces are then brought to the same temperature by filling the box with hot water. When the cube thus filled is placed before a thermo-pile, the indications of the latter instrument show at once the relative amounts of energy which the various faces of the cube emit by radiation.

It will then be seen that the polished surface radiates less than the roughened one, the latter far less than the one coated with lamp-black, while the painted one will radiate comparatively little if the color is light, especially if it is white, but more and more in proportion as the color is darker, and nearly as much as lamp-black if the pigment is black. The polished surface, which is thus seen to be a poor radiator, is a good reflector and therefore a poor absorber, while the blackened surface is a good radiator, but it is a poor reflector and hence a good absorber.

CHAPTER XVII.

SPECIFIC HEAT.

239. The Unit Quantity of Heat—The Calorie. In making thermal measurements, a clear distinction must be drawn between temperature and quantity of heat. Temperature, as we have seen, (p. 234), is a measure of the kinetic energy of the molecules of a body, but is entirely independent of the number of these molecules, that is, of the mass of the body. On the other hand, the quantity of heat a body contains denotes the total kinetic energy of its particles, and this depends upon the mass of the body as well as upon the velocity with which these particles are moving.

Since it is clear that a perfectly definite amount of heat will raise a certain quantity of a particular substance through a given range of temperature, such a quantity of heat may be taken as the unit quantity. It is found, however, that the quantity of heat required to raise the temperature of a given mass of one substance to a certain degree will raise the temperature of the same mass of another substance to an entirely different degree. Therefore, in choosing the unit quantity of heat, it is necessary to specify the substance, as well as its mass and the range of temperature through which it is to be heated. The substance thus selected is water, its mass is one kilogram, and the range of temperature through which it is heated is from 19.5° C. to 20.5° C. *The unit quantity of heat, therefore, is the quantity of heat that will raise the temperature of* 1 *kilogram of water from* 19.5° C., *to* 20.5° C., *and is called the calorie.* Inasmuch, unfortunately, as another unit of heat is sometimes employed where one gram is used in place of one kilogram, it is necessary to distinguish these units of heat as *the greater caloric* and *the lesser caloric*. The lessor calorie is 1000 times smaller than the greater.

240. The Choice of the Calorie. Experiment shows that the quantity of heat required to raise the temperature of a kilogram of water from 9° C. to 10° C. is slightly different from that employed where the range of temperature is from 19° C. to 20° C. Indeed, a similar variation of greater or less amount is found when any other two successive degrees are thus compared throughout the whole thermometric range. Accordingly in determining the calorie it is necessary to specify the particular degree through which the rise of temperature of the kilogram of water is to be carried; and while other calories are likewise in use for special purposes, the one here given, which is known as the 20° calorie, possesses certain advantages, one being the fact that this temperature is approximately the temperature of the room in which observations are usually conducted. The variation, however, in the quantity of heat required to raise the temperature of a kilogram of water 1° C. is not large throughout the entire range of temperature from 0° C. to 100° C.; and hence, except in measurements of extreme accuracy, we may, for all practical purposes, take the calorie to be the quantity of heat required to raise the temperature of 1 kilogram of water through 1° C. Roughly, therefore, it requires 100 calories of heat to raise the temperature of 1 kilogram of water from 0° C. to 100° C.

241. Differences in Thermal Capacity. When equal quantities of heat are applied to equal masses of several different substances, it is found, in general, that the resulting temperatures of these substances are widely different. On the other hand, if equal masses of several different substances are raised to the same temperature, each substance will absorb a quantity of heat different in amount from that taken up by any other substance. The quantity of heat, for instance, that will raise the temperature of a kilogram of water from 0° C. to the boiling point will raise the temperature of a kilogram of iron to a bright red heat. This shows that more heat is required to raise a mass of water to a given temperature than to raise an equal mass of iron to the same temperature. Accordingly, substances differ from one another in *their capacities for heat;* and the substance that requires the largest amount of heat, mass for mass, to produce a given rise of temperature has the largest capacity for heat. Thus, in the illustration just given, water has a capacity for heat much greater than iron, and, indeed, greater than that of any other known substance except hydrogen.

242. Specific Heat. When the mass of the substance is 1 kilogram and the range of temperature through which it is heated is 1° C., the quantity of heat required to effect this change of temperature is known as the specific heat. *The specific heat of a substance, therefore, is the number of calories of heat required to raise the temperature of 1 kilogram of the substance through 1° C.* From this it follows that the specific heats of substances are in the same ratio as their thermal capacities, a body which has a high specific heat requiring a relatively large amount of heat to raise it to a given temperature.

Experiment shows that the specific heat of a substance is not always the same, but depends upon the circumstances under which the determination is made. The specific heats of solids and liquids are slightly different at different temperatures, while the specific heats of gases show marked deviations when the pressure and volume conditions of the gas undergo change. So slight, however, are the variations in the specific heats of solids and liquids, that for approximate measurements they may be neglected. With gases, on the other hand, it is entirely different: and in all measurements in which the specific heats of gases are involved, the most careful attention must be paid to the conditions of pressure and volume when heating takes place.

243. Causes of the Differences Between Specific Heats. The question as to why one substance requires more heat than another to produce a given rise of temperature may be answered by considering the effects of heat upon a body.

If the body, when heated, expands, and so occupies a larger volume, the atmospheric pressure, together with any other pressure to which it is subjected, must be pushed back. This force, acting through the distance through which the body expands, performs work; and the energy which is expended in performing this work must, of course, be derived from the heat. A portion of the energy of the heat is thus consumed in performing work exterior to the body, that is, in doing *external work*.

Another portion, and often a very large portion, of the heat supplied to a body is expended in performing work inside the body itself, or in doing *internal work*, and this may be accomplished in several different ways. When the kinetic energy of the molecules is increased with a rise of temperature (p. 234), the blows which they deliver upon one another become more frequent and more violent, so that the molecules are crowded farther apart (p. 243), and the body expands. In all cases of expansion, therefore, the molecular, or cohesive, forces are moved through a certain distance, and this requires the expenditure of work. Obviously the energy necessary to perform this work, which has its equivalent in the resulting increased potential energy of the molecules, is drawn likewise from the supply of heat.

Experiment indicates, moreover, that a portion of the energy of the heat is expended in giving the atoms within the molecules not only an increased rotary or vibratory motion among themselves about their centers of mass (p. 96), but also an additional energy of position, or potential energy, with respect to one another. Finally there are reasons to believe that the electrons within the atoms (p. 10) are accelerated in their orbital rotations by a rise of temperature, and this also involves a consumption of energy. In all of these ways, energy is consumed within the body itself, and thus may be said to do internal work.

When heat is applied to a body, therefore, this heat is expended partly in doing external work, partly in doing internal work, and partly in raising the temperature. The amount available for raising the temperature is only that which remains after the performance of the external work and the various forms of internal work. If expansion is prevented while the body is being heated, it is evident that no external work can be performed.

In this case a larger proportion of the energy is then available for raising the tempeature, and the body becomes hotter. Again, if the conditions are such that the major part of the energy supplied is employed in doing internal work, but little is left for raising the temperature and effecting the consequent expansion.

244. The Two Specific Heats of a Gas. When heat is added to a gas to raise its temperature, there prove to be two important conditions under which this application of heat may be made: first, when the gas is confined in a tight vessel which does not allow it to expand, and which, therefore, maintains the volume constant; second, when it is enclosed in a cylinder provided with a piston which moves freely, so that the gas, as it expands, remains at constant pressure.

In the first instance, when the gas is kept at constant volume, no external work can be done, since no expansion occurs. The effect of the heat then, is confined solely to increasing the kinetic energy of the molecules, that is, in raising its temperature and to doing internal work upon the atoms and electrons within the molecules. If the vessel contains 1 kilogram of the gas, the number of calories thus consumed in raising the temperature 1° C. is the specific heat of the gas at constant volume, and is usually denoted by C_v.

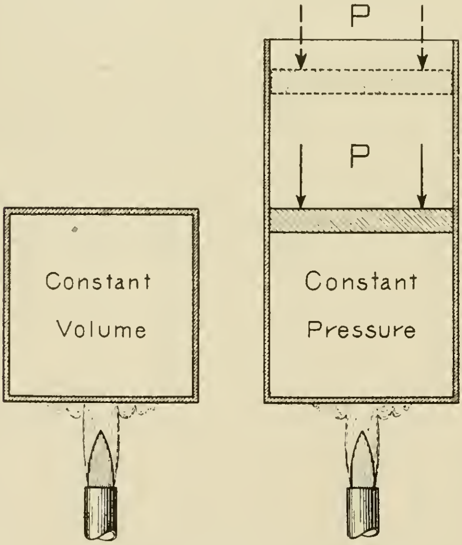

FIG. 166. The specific heat of a gas at constant pressure is greater than that at constant volume.

In the second instance, when the gas is heated in a cylinder provided with a freely-moving piston, Fig. 166, the pressure remains constant throughout, and the gas in expanding, pushes back the piston, together with the pressure upon it, and so does external work. To the energy which is consumed thus in raising the temperature and in doing external work, must be added a minute amount expended in doing internal work upon the molecules of the gas. The slight forces of attraction which the molecules of a gas have for one another are moved through a certain distance when the molecules are separated, as the gas expands, and so internal work is done between the molecules. This work increases the potential energy of the molecules, and therefore increases the quantity known as *the inner energy of the gas.* The latter energy, however, is relatively small in amount, and, except in measurements of great precision, may be neglected.

Practically, then, the energy which is supplied to heat a gas at constant pressure is expended in raising the temperature of the gas and in doing external work. If the cylinder contains 1 kilogram of the gas, and this is heated 1° C., the amount of heat thus supplied is known as the specific heat at constant pressure, and is denoted by C_p.

The ratio of these two specific heats, which increases with the pressure, is constant for all diatomic gases at atmospheric pressure, and has the value 1.41. That is, for diatomic gases at atmospheric pressure

$$\frac{C_p}{C_v} = 1.41.$$

The specific heat of a gas at constant pressure may be easily determined by experiment, but until recently the specific heat at constant volume could be neither so readily nor so accurately measured. Previously the ratio between them, that is,

$$\frac{C_p}{C_v} = 1.41.$$

could be determined only from the velocity of sound in the air, by a method to be described later; and as C_p was known by actual measurement, C_v could be accurately determined.

It is interesting to note that, although the ratio of the two specific heats is 1.41 for all *diatomic* gases, this same ratio has the value 1.66 for *monatomic* gases, such as argon, helium, and mercury vapor, while it falls as low as 1.025 for a *polatomic* molecule, such as ether vapor, which has fifteen atoms in its molecule. The explanation is simple when we consider that no internal work is possible within a molecule that consists of one atom only, the atom and the molecule in this case being identical. In a molecule which contains two atoms, a certain portion of the energy is consumed in doing work upon these atoms, that is in giving them increased kinetic energy as well as a greater potential energy with respect to each other. As might be expected, this effect is multiplied in a molecule containing many atoms; for here the work done within the molecule is very great indeed, as compared with the external work which the gas does in expanding.

245. Atomic Heat—Dulong and Petit's Law. A remarkable relation, discovered by Dulong and Petit, exists between the specific heats of certain substances and their atomic weights. These investigators found that, with very few exceptions, the specific heats of all solid elementary substances are inversely proportional to their atomic weights, or that the product of the specific heat and the atomic weight of all solid elementary substances is constant, and equal approximately to 6.4. This remarkable law has been of the greatest service to the chemist as an aid in determining atomic weights, and points to a most interesting fact regarding the capacities of the atoms for heat.

Since the relative weights of the atoms are expressed by their atomic weights, the actual weights in grams of an equal number of atoms of different substances are to each other as their atomic weights. That is, if as many grams of different substances are taken as there are units in their atomic weights, these masses contain the same number of atoms. Thus, since the atomic weight of iron is 56, that of copper 63, and that of lead 208, it follows that there are as many atoms in 56 grams of iron as in 63 grams of copper or 208 grams of lead. If we multiply these quantities by the specific heats of these substances, we shall obtain 6.4, which is the number of lesser calories required to raise the temperature of 56 grams of iron, 63 grams of copper, or 208 grams of lead 1° C., that is, to raise the temperature of the same number of atoms of iron, copper, and lead through this range of temperature. This product of the atomic weight and the specific heat of a substance is known as *the atomic heat* of that substance.

From the remarkable fact that the atomic heat of nearly all solid elementary substances is constant, we may draw the following conclusions:

1. The quantity of heat that will raise the temperature of one atom a certain amount will also raise the temperature of any other atom by the same amount; or, the capacity for heat of one atom is the same as that of any other atom, and is wholly independent of the nature of the substance.

2. Masses of different substances proportional to their atomic weights are said to be chemically equivalent, that is, one may replace the other in any chemical compound. Dulong and Petit's law states that equal quantities of heat are required to raise chemically equivalent substances to equal temperatures, or in other words, substances that are chemically equivalent are also thermally equivalent.

The law of Dulong and Petit is true for nearly all elementary bodies in the solid state, but holds only imperfectly for gases and not at all for liquids. The law, however, has been extended by Kopp so as to apply also to solid compound bodies. With these it is found that the *molecular heat* of a body, that is, the product of the molecular weight and the specific heat, is equal to the sum of the atomic heats of the elements therein contained.

CHAPTER XVIII.

FUSION AND SOLIDIFICATION.

246. Fusion. The molecules of a solid must be regarded as having a definite position with reference to one another, and as being held in that position by the strong molecular forces that bind molecules together. The molecules of a solid body, therefore, are in a position of stable equilibrium under the action of these molecular forces.

When heat is applied to a solid, the kinetic energy of its particles increases until a perfectly definite temperature for each substance is reached, after which the rise of temperature ceases altogether. A further addition of heat, instead of producing a corresponding increase of temperature, causes the strong molecular bonds to be broken and the molecules to be forced farther apart, in opposition to these molecular forces that hold them together. When the molecules are no longer held so rigidly together, they are free to leave the molecules to which they have been bound, and to wander among other molecules, and the body passes from the solid to the liquid state, or, in other words, the body melts. When the molecules are thus separated from one another, work must be done in overcoming these molecular forces; hence the passing from the solid to the liquid state requires the consumption of energy, and this energy is supplied in the form of the heat which is necessary to melt the substance. Thus, when ice passes from the solid to the liquid state, the heat required to melt the ice is consumed in tearing apart the molecules and thus giving to the molecules of water a potential energy which they did not possess as ice.

247. The Heat of Fusion. The heat thus consumed during fusion is expended, not in raising the temperature, but *in doing internal work, that is, in giving the molecules a potential energy which they did not before possess.* Although a large amount of heat is expended in the process of fusion, the temperature of the resulting liquid is absolutely the same as that of the solid before fusion. In the melting of ice, it is found that 80 calories of heat are required to change a kilogram of ice at 0° C. to water at 0° C. A kilogram of water at 0° C. has, therefore, more energy than a kilogram of ice at 0° C. by an amount equivalent to 80 calories of heat. The heat which thus disappears during fusion, and which is known as *the heat of fusion,* is to be considered as consumed in giving the molecules of the liquid a potential energy with reference to one another, and must never be regarded as "latent".

248. The Laws of Fusion. The laws according to which fusion takes place are three in number, and are as follows:

1. Every crystalline substance has a perfectly definite melting point, which, under ordinary conditions, is the same as the freezing point.
2. When the melting point is reached, the substance remains at a constant temperature until it is entirely melted.
3. Every substance requires a perfectly definite quantity of heat to change it from the solid to the liquid state.

249. Freezing Mixtures. The efficiency of a freezing mixture, such as ice and salt, depends upon the fact that heat is required to melt a substance. The chemical action between salt and ice is such that the two cannot remain together in the solid state; the salt causes the ice to melt, and the water thus produced dissolves the salt. Every kilogram of ice in melting consumes 80 calories of heat; and as no other source of heat is present, the energy required to melt the ice is taken from the ice itself. As heat, therefore, is withdrawn from ice at 0° C., the temperature of the mixture falls far below 0° C.

250. Solidification. When a liquid solidifies, as, for instance, when water freezes, the above process takes place in the reverse order, and the liquid falls in temperature until the freezing point is reached. The freezing point, which, in general, is the same as the melting point, is the temperature at which liquid particles come together and form a solid. While the body remains a liquid, the kinetic energy of the particles is too great to permit of a permanent union between the molecules, so that the latter remain separated. their molecular forces being constantly disrupted. As the temperature falls, however, the molecular bombardments decrease in violence, until, when they are no longer able to prevent the molecules from uniting, the strong molecular forces draw these particles together, and the liquid solidifies. The molecular forces, thus moving through a distance, do work upon the molecules, and give to them an added kinetic energy equal in amount to the potential energy which they possessed during the entire time the body had been liquid.

251. The Heat of Solidification, Accordingly, as the liquid solidifies and the particles come again into the position of maximum stability, the potential energy which the molecules possessed in the liquid state is transformed into kinetic energy. In other words, the heat which had previously been required to melt the solid, and which had remained in the liquid, not in the form of heat, but as potential energy of the molecules, now becomes kinetic energy once more, and warms the body. The heat which is thus evolved when a liquid solidifies is known as *the heat of solidification*, and is identical in amount with the heat of fusion, which had at some previous time been applied to melt the solid. It is thus seen that the energy represented by the heat of fusion remains within the liquid as molecular potential energy until it is transformed into molecular kinetic energy, or the heat of solidification, which becomes apparent as heat when the liquid passes into the solid condition. For every kilogram of ice that freezes, 80 calories of heat are thus "set free," and this heat raises the temperature of the water and the adjacent air.

252. Supercooling and Supersaturation. By taking great precautions to exclude the air and to prevent agitation, it is possible to cool a mass of water even 16° C. below zero before it solidifies. In this condition, which is known as that of supercooling, the particles are in a state of unstable equilibrium, and the water will instantly pass into the solid state as soon as freezing begins. A slight jar is sufficient to overturn these insecure molecules, hence a mass of supercooled water will partly solidify at once when it is in any way agitated, or when it is merely touched with a crystal of ice, while the temperature will immediately rise to 0° C. because of the energy evolved as the heat of solidification.

Analogous to the condition of supercooling is that of the supersaturation of a salt. A salt, such as sodium acetate, which may be dissolved by heating with a very little water, if allowed to cool in a quiet place, will cool below the temperature at which it normally crystallizes without solidification taking place. The molecules of this solution are, however, in a thoroughly unstable condition and will pass into the solid state as soon as crystallization begins. The overturning of this unstable equilibrium may be started by dropping into the solution the smallest fragment of a crystal of the solid salt. Crystallization then proceeds rapidly throughout the liquid, and the potential energy possessed by the particles in the liquid form now becomes kinetic, and the solid rises many degrees in temperature. This heat which is thus set free as the heat of solidification when a liquid passes into the solid state, is the same in amount as the heat that was consumed as the heat of fusion when the solid melted.

253. Expansion upon Freezing—The Crystalline Structure of Ice. The fact that water expands as it freezes—the volume of the ice being about one-eleventh greater than that of the water before freezing took place—would seem to indicate an error in the statement that the molecules of ice are closer together than the molecules of water. The explanation of this apparent contradiction lies in the fact that ice is a crystalline body, and that the molecules are joined to one another, as it were, by their corners, or poles. This may be easily understood if we assume that the molecules are angular instead of round, and that the points of attachment between them are situated in these angles. In the liquid state, while the kinetic energy of the molecules prevents a union of the poles, we may conceive the angles as interlocking, or fitting into one another, like the cells in a honeycomb, Fig. 167, so that the liquid occupies a relatively small volume. On the other hand, when solidification occurs, that is, when a polar union between the molecules takes place, the molecules can join one another, two and two, only by their angular extremities, Fig. 168, and consequently an increase of volume must necessarily accompany the decrease of potential energy which thereby results. The illustrations here given, which attempt to show how such an arrangement could exist, must be regarded, however, as purely diagrammatic and not at all as representing the actual molecular condition.

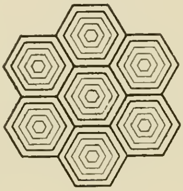
Fig. 167. Possible arrangement of the molecules in water.

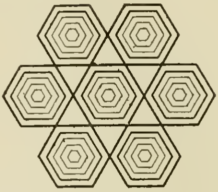
Fig. 168. Possible arrangement of the molecules in ice.

As a matter of fact, it is possible and, indeed, highly probable that the water molecule is actually hexagonal in its polar attraction; for when a ray of sunlight, or the beam from an electric lantern, is sent through a block of ice and the particles are examined with a microscope, it will be seen that the ice is not a confused mass of solidified water particles, but consists of regular hexagonal crystalline structures of remarkable symmetry and beauty, Figs. 169 and 170. These beautiful forms, which

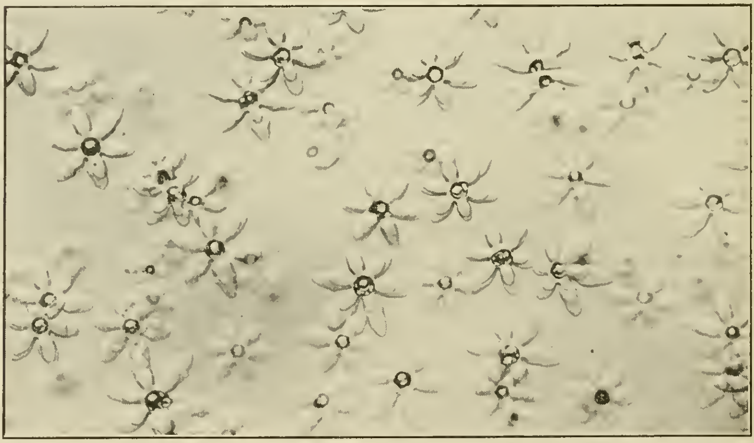
Fig. 169. Ice flowers.

are known as the ice flowers, are individual crystals of ice which have been melted by the heat of the lantern, and now exist as water masses of the same delicate shape within the ice. The globule at the center of each crystal does not appear until the crystal melts, and is not, as might be supposed, a bubble of air, but is a nearly perfect vacuum. The water produced by the melted crystal must occupy a volume one eleventh smaller than the crystal itself; and this vacuous space, filled only with the vapor of water, thus represents the difference between the volume

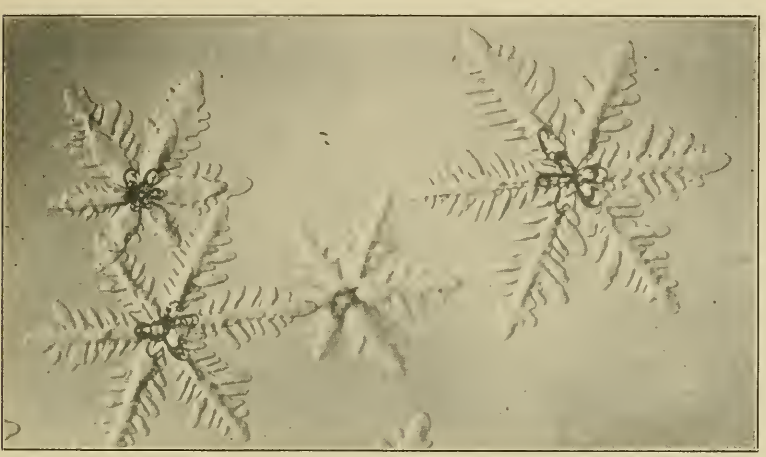

FIG. 170. Ice flowers.

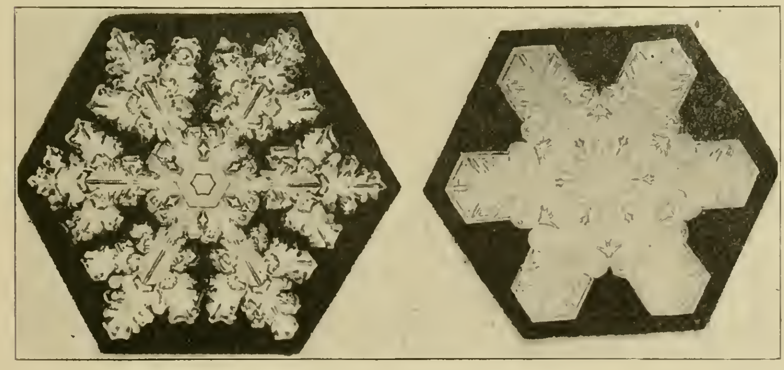

FIG. 171. Snow crystals.

of the ice crystal before melting and that of the resulting water. Even greater regularity of formation is observed in crystals of snow, Fig. 171, where the exterior beauty of design is excelled only by the exquisite tracery of the interior.

The molecular forces called into play in forming these delicate crystalline structures are enormously great, as may be seen in the bursting of frozen water pipes. The molecules of water can not pass into the solid, crystalline state without filling a volume one-eleventh greater than that which they occupied as a liquid, and so the strongest pipes are broken when the water within them freezes. Similarly, a hollow cast-iron sphere, with walls a centimeter or more in thickness, will be broken in pieces when the spoonful of water that fills its interior cavity is frozen.

254. The Effect of Pressure upon the Melting Point. Substances, like ice, that contract upon melting, have their melting points lowered very slightly by pressure. Melting consists not only in giving freedom of motion to the particles, but of causing the particles to occupy a volume smaller in the liquid state than in the solid. The application of pressure to a substance like ice tends to compress the molecules into smaller volume, and indeed does reduce the volume to a very slight extent. Pressure, therefore, facilitates the melting since it tends to reduce the volume, and hence it slightly lowers the melting point. It is found that for each atmosphere of added pressure, the melting point of ice is lowered 0.0074° C., or $\frac{1}{135}$° C.; and accordingly, under a pressure of 135 atmospheres, the ice could melt at a temperature of —1° C.

Just as an increase of pressure lowers the melting point of ice, so also does a decrease of pressure slightly raise this point. Accordingly, ice, which melts at 0° C. under a pressure of 1 atmosphere, will not melt until its temperature is raised to + 0.0074° C. if the pressure is reduced to zero, that is, if the melting takes place in a vacuum. At the pressure of 4.6 millimeters of mercury, which occurs in an experiment to be described later, the melting point of ice is + 0.0073° C.

This lowering of the melting point of ice by great pressure may be illustrated by suspending a heavy weight from a wire thrown across a block of ice. If the wire is sufficiently fine, the entire pressure due to the weight of the suspended mass is brought to bear upon the small surface in contact with the ice. This may amount to as much as 100 or 150 kilograms per square centimeter, and a pressure as great as 135 atmospheres may therefore be exerted between the wire and the ice. In this case the ice immediately beneath the wire would be able to melt at —1° C. because of the great pressure to which it is subjected, but being at 0° C., it is above its melting point, and therefore melts. The water produced by the melting ice flows behind the wire and there freezes. Accordingly, with melting taking place before and freezing taking place behind, the wire in this way passes entirely through the block, leaving the ice as solid as in its original condition.

255. Regelation. If fragments of ice are placed in a mould and subjected to great pressure, the melting point is lowered by the pressure, and a partial melting takes place. When the pressure is relieved, the liquefied ice solidifies immediately, so that it is thus possible by pressure to mould fragments of ice into a clear, crystalline mass. The process of forming a mass of clear, solid ice from a confused collection of fragments, due to the fact that liquefaction occurs when pressure is applied, is called regelation. In making a snowball, the pressure of the hands cause a partial melting of the snow crystals to take place at the angles and edges where they come in contact with one another, and these semifluid crystals then solidify into a more or less solid mass as the pressure is relieved. In the same way, two pieces of ice may be welded by pressing their edges tightly together, even though this be done under warm water.

A result observed in the bursting of frozen water pipes beautifully illustrates this lowering of the freezing point by pressure. Freezing cannot take place within the pipe until the water occupies a larger volume, and this is prevented by the rigidity of the metal. An enormous pressure is therefore developed and the temperature of the water falls, possibly several degrees below 0° C., remaining all the while perfectly fluid until the pipe is broken. When the rupture is effected, freezing takes place at once, and a tongue of ice protrudes from the opening. The ice thus thrust out shows that the water had remained liquid until the instant of bursting, and that the enormous pressure within the pipe had lowered the freezing point probably many degrees. The solidification, which was thereby retarded, could not take place until the bursting of the pipe enabled the water, in becoming ice, to occupy a larger volume, and then the change to the solid state was accomplished instantly.

256. The Glacier. The lowering of the melting point of ice due to pressure is illustrated on a grand scale in nature by the movement of the glacier in its slow, creeping progress down the valley. The accumulated snows of many years, partially liquefied by the enormous pressure which the lower layers must sustain, finally congeal into a mass of solid ice, having a peculiar granular structure characteristic of the glacier. Each individual crystal of this ice must sustain the weight of the overlying masses, and the edge, or angle, where one crystal meets another is really the point at which this pressure is brought to bear. Here the melting point is lowered by the pressure, and a minute portion of the ice therefore melts. The water produced by this liquefied ice, since it is acted upon by gravity, flows to other adjacent and, in general, lower parts, where the pressure is less, and there it freezes. The pressure, which is thereby relieved at this particular point, now develops elsewhere, and here the same process is repeated with the same result.

The downward movement of the glacier is therefore caused by the momentary melting of minute particles of the ice, followed by an immediate re-congealing of the liquefied portions as soon as these are urged by gravity and pressure to regions where the pressure is less. Accordingly, in this way, by a movement which is infinitessimal for any one particle, but which is continuous throughout the entire mass of ice, the glacier advances down the valley often at the rate of several feet a day, winding in and out between the hills on either side, and having the appearance of a slowly-flowing river of a very viscous liquid.

CHAPTER XIX.

VAPORIZATION AND CONDENSATION.

257. Evaporation. When the molecular collisions take place upon the interior of a body of fluid, one particle in encountering another, may give to that other a portion of its energy, and have its own kinetic energy decreased. Inasmuch, as the second particle gains in energy as much as the first one loses, the total kinetic energy, and therefore the temperature of the fluid, remains unchanged.

This, however, is not the case when these molecular bombardments take place at the surface of a liquid. A particle at the surface may be struck by a particle directly beneath, and if the blow is sufficiently great, this surface particle may be projected to such a distance from the body of the liquid that the bonds that bind together the molecules of a liquid may be broken, and this little surface particle may now wander about in the space above the liquid. The same process may be repeated with other particles, and soon the space above the surface is filled with molecules that were formerly a part of the body of the liquid, but now are separated so far from one another that there is only an extremely weak molecular force acting between them, and they now exist in the state of a gas or vapor. This process is known as evaporation.

Evaporation, therefore, is the gradual process of passing from the liquid to the gaseous state by the projection of particles from the liquid surface into the space above the surface. It is to be observed that the particles thus torn away must possess a kinetic energy greater than that of the average particle, since they carry with them a portion of the energy of the particles that collide with them. That is, the interior particles, thus colliding with those on the surface, give to the latter a portion of their own energy, while they themselves remain in the liquid. The particles leaving the surface therefore carry with them kinetic energy which formerly belonged to those that still remain in the liquid, hence the liquid falls in temperature. This explains why it is that evaporation is a cooling process. If external heat is applied to the liquid, whereby its temperature is raised, the kinetic energy of its molecules is increased, a larger number of particles will then be projected from the liquid surface in a given time, and evaporation takes place more rapidly at the higher temperature.

Since the molecules of all bodies are in motion, many solids undergo a very slow process of evaporation so, that even a body like ice very gradually evaporates. The clothes on the line are often said to "freeze dry."

258. Vapor Tension—Boiling. Nearly all liquids in contact with the atmosphere dissolve the air to a greater or less extent, so that frequently air in large quantities exists in the form of tiny, microscopic bubbles distributed through the body of the liquid. If the process of evaporation takes place at the surface of these bubbles, Fig. 172, the interior of a little bubble is then filled with particles of vapor that collide with one another and strike against its liquid walls. These impacts against the walls of the little bubble produce a pressure which is known as *the vapor tension.*

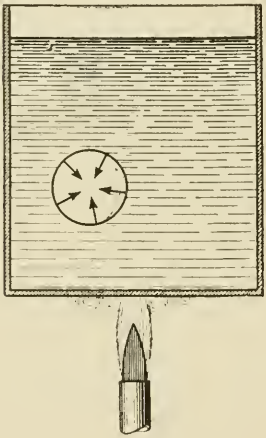

Fig. 172. The cause of vapor tension.

The vapor tension becomes greater as the temperature of the liquid is raised, and the kinetic energy of the particles of the vapor in the bubble is increased. At a certain temperature the pressure within the little bubble becomes equal to the atmospheric pressure without, the bubble then increases enormously in size, and, in rising to the surface, produces the phenomenon known as boiling. Boiling, therefore, is the phenomenon of evaporation taking place beneath the surface of a liquid, whereby bubbles of the vapor rise through the liquid to the surface.

Boiling, however, cannot take place until the temperature rises to such a degree that the pressure of the vapor, or the vapor tension, at that temperature, is equal to the external pressure; for at a lower temperature the vapor tension is less, hence the external pressure crushes the little bubble of vapor about to form, and prevents it from rising to the surface. Accordingly, the boiling point of a liquid may be defined as the temperature at which the vapor tension of the liquid is equal to the atmospheric pressure. An increase of external pressure raises the boiling point, while a decrease of pressure lowers it.

The lowering of the boiling point due to a decrease of pressure may be illustrated by removing from the flame a flask of boiling water and corking it immediately, after which the pressure of the vapor upon the surface of the water quickly prevents further boiling. If cold water is now poured upon the flask, the vapor will be condensed, and, the pressure being thus diminished, the water boils furiously. By thus condensing the vapor by pouring on cold water, the water within the flask may be made to boil until it falls to a temperature but little above that of the room.

259. Explosive Boiling—Bumping. It has been seen that the process of evaporation may take place from a liquid into the little air bubbles dissolved in the liquid as well as from the free upper surface. The particles of vapor, as they fly about within this little bubble, exert a pressure, or tension, upon its sides, so that the bubble tends to expand. If, however, the external pressure is greater than the vapor tension of the liquid at that temperature, the bubble is prevented from increasing in size and rising to the surface. As soon, however, as the temperature rises to the point at which the vapor tension is equal to the external pressure upon the liquid, the little bubble increases in size, more liquid evaporates into it, the bubble of vapor rises through the liquid, and the liquid boils.

If a liquid that does not contain dissolved air is used, or if water is boiled until all its dissolved air has been carried away with the bubbles of vapor, that is, until all the air has been "boiled out," there is then no little cavity to act as the nucleus of a bubble of vapor, into which the particles of the vapor may be projected, and as a consequence no bubble of vapor forms. The energy supplied by the source of heat then remains in the water, and the temperature rises far above the boiling point. The vapor tension increases at the same time, until finally, when from some cause a bubble of vapor does form, vapor will be projected into this bubble with great rapidity, and the water boils with explosive violence. In this way boiler explosions are often explained.

This violent production of vapor is observed in a vessel of water that has been boiling for some time, and produces the phenomenon known as "bumping." If a piece of pumice stone is dropped into the water when this effect occurs, the air contained in the pumice stone furnishes centers about which bubbles of vapor may form, and the boiling becomes normal at once. A small piece of platinum wire, which absorbs large quantities of air, produces the same effect.

260. The Heat of Vaporization. When heat is applied to a liquid, the temperature rises until the boiling point for that pressure is reached, and at this point the temperature remains perfectly constant until the liquid is entirely boiled away. If more heat is supplied, the liquid boils harder, but the temperature remains constant. After the boiling point is reached, part of the energy supplied to the liquid is consumed in doing internal work in separating the molecules from one another as they leave the liquid to form the vapor, and part is expended in doing external work in pushing back the atmosphere when the smaller liquid volume changes to the larger gaseous one. Under normal conditions steam occupies a volume about 1,700 times as great as the volume of the water from which it was produced. Hence the heat supplied to water after boiling has begun is consumed partly in doing internal work, that is, in tearing apart the molecules of water and giving them a potential energy as steam which they did not possess as water, and partly also in doing external work in pushing back the atmosphere to make room for this very greatly increased volume of vapor.

The molecular bonds which must here be ruptured are very strong indeed, and a large amount of heat must be applied to change water from water at 100° C., to steam at 100° C. Indeed, for every kilogram of water passing from water at 100° C., to steam at 100° C., 537 calories of heat must be supplied. These 537 calories of heat are not consumed in raising the temperature, for the temperature of the resulting steam is no higher than that of the boiling water. This large amount of heat exists in the steam in the form of the potential energy of its particles, so that the energy in a kilogram of steam at 100° C. is greater than that in a kilogram of water at the same temperature by an amount equal to 537 calories. The heat per kilogram of liquid which thus disappears in the passage from the liquid into the gaseous state, and which is called *the heat of vaporization,* must on no account be called "latent", for it is in no way hidden within the liquid. It has, indeed, been expended in doing external and internal work, and still exists, not as heat, but as potential energy, in the particles of the vapor.

261. The Laws of Ebulition. The laws according to which boiling takes place are five in number, as follows:

1. Every liquid boils at a temperature which is constant for that liquid under similar conditions and pressure.
2. An increase of pressure raises the boiling point, and a decrease of pressure lowers it.
3. The boiling point is raised by salts and other solids dissolved in the liquid.
4. After boiling begins, the temperature remains perfectly constant until the liquid is entirely boiled away.
5. A definite quantity of heat per kilogram of liquid is required to change a liquid into the state of vapor.

262. The Heat of Condensation. It is most interesting to examine the converse of the process of evaporation, and to consider the phenomenon of the condensation of a vapor. If a quantity of steam is cooled, that is, if energy in the form of heat is withdrawn from steam at 100° C., the particles are then able to approach so near to one another that the molecular forces that bind particle to particle and form a liquid can act. The particles of steam, hitherto separated from one another, now rush together, forming the more stable compound water, and the potential energy possessed by the molecules as steam now becomes kinetic, and is consumed in warming the remainder of the steam, and the surrounding air. For every kilogram of steam thus condensed from steam at 100° C. to water at 100° C., 537 calories of heat are therefore "set free", that is, this amount of energy passes from the form of the potential energy of the particles of steam, to the form of the kinetic energy of the particles of the surrounding objects which are thereby raised in temperature.

It is by means of this heat, which is called *the heat of condensation*, that modern steam-heated buildings are warmed. The water condensed in the radiator flows back through the same pipe that carries the steam, hence both water and steam are at the same temperature. When the steam condenses into water, the potential energy of its molecules is transformed into the kinetic energy of the particles of the radiator, 537 calories of heat being thus communicated to the radiator for every kilogram of steam that condenses, and this is the heat that is effective in warming the building. The heat of condensation, therefore, which now appears as molecular kinetic energy, is thus identical in amount with the heat of vaporization which was added to the water to volatilize it, and which has remained in the steam, not as heat, but in the form of the added potential energy of the molecules.

263. The Freezing of Water by Its Own Evaporation. A most interesting experiment, illustrative of the fact that evaporation is a cooling process, is the freezing of water by its own evaporation. We have seen, (p. 276), that a liquid will boil at any temperature, that is, that bubbles of vapor will form within the liquid and will rise to the surface, when the pressure upon the liquid does not exceed the tension, or pressure, which the vapor exerts in the tiny air bubbles dissolved in the liquid. Accordingly, if a small quantity of water is placed under the receiver of an air pump and the exhaustion of the air is begun, the pressure soon falls to a value equal to the pressure of the water vapor at that temperature, and the water boils. Each particle of the vapor, however, as it leaves the water surface, carries with it a portion of the energy which the water originally contained; and, as all flow of heat to the liquid is carefully eliminated, the water, as it boils, loses in energy and rapidly falls in temperature.

If the pressure is reduced as low as 4.6 millimeters of mercury, this process of boiling and consequent cooling continues until the temperature + 0.0073° C. is reached, which (p. 272) is the temperature at which ice melts at this low pressure, and therefore the temperature at which water freezes under this same pressure. In other words, at the pressure of 4.6 millimeters of mercury and at the temperature of + 0.0073° C., the boiling point and the freezing point of water fall together. The slightest reduction in pressure below this amount will cause the water to boil; and, as this boiling water is also at its freezing point, the further loss of energy due to the boiling will cause it likewise to freeze, and the water freezes even while it boils.

In order that water may boil at the temperature at which it freezes, it is obvious that the external pressure upon it must not be greater than the vapor tension of the water at this temperature, which is 4.6 millimeters of mercury. In a vacuum, however, the evaporation is so great that the vapor is formed more rapidly than it can be removed by the pump, so that the pressure of the vapor soon prevents further boiling. It is therefore necessary to remove this vapor by other means, and this may best be accomplished by absorbing it with sulphuric acid. Accordingly, the capsule, or watch glass, containing the water to be frozen is supported immediately above a broad, flat dish containing strong sulphuric acid. The acid, by its affinity for water, absorbs the water vapor as quickly as formed, thus leaving the air to be removed by the pump.

As the exhaustion of the air proceeds, continued boiling, with consequent cooling, takes place, until, at the pressure of 4.6 millimeters of mercury, the temperature + 0.0073° C. is reached. This is the pressure and this the temperature at which water, water vapor, and ice can exist in equilibrium with one another: that is, at this point no one of these three states in which water can exist will increase at the expense of the others, so that this point is known as the triple point. At a pressure below that of the triple point, boiling and freezing proceed at the same time, until the water is entirely solidified.

264. The Geyser. The action of the geyser is a beautiful illustration of the principle that the boiling point of a liquid increases with the pressure. The geyser consists of a cavity in the earth filled with water and heated at the bottom. At the surface the water is at atmospheric pressure, and at each point within the geyser the pressure is greater than one atmosphere by the weight due to the column of water above it. The pressure upon each particle of water within the geyser, therefore, is proportional to the depth of that particle below the surface. Since the boiling point depends upon the pressure, the boiling point is higher at points far down within the geyser than at points nearer the surface; that is, the deeper the point within the geyser, the higher must be the temperature of the water at that point before boiling can begin.

Accordingly, if the water throughout the geyser had a uniform temperature of 100° C., boiling would take place at the surface, but could not take place in the interior. The internal heat of the earth, therefore, raises the temperature of the water in the geyser until the water at the surface is 100° C., and that below the surface is at a temperature higher than 100° C. in proportion to the depth of the water below the surface. That is, the water at each point within the geyser is heated to the temperature just below that at which it can boil under the pressure existing at that point. Accordingly, at the instant before the eruption takes place, every particle of water within the entire column is just on the point of boiling, but it cannot boil because the vapor pressure at any point is just a little less than the total pressure to which the water at that point is subjected. A moment later, however, more heat passes into the column of water from below, thus increasing the vapor tension of the water at the bottom to such an extent that this is equal to the pressure already existing at the bottom, and therefore a bubble of vapor forms.

This bubble of vapor, forming at the bottom of the geyser, lifts the entire column of water within the tube. A little water runs over at the top, and this causes the pressure upon every particle of water within the column to be less than it was before. The water at each point, however, was just at its boiling point at the pressure under which it existed. The pressure now being somewhat lessened, a bubble of vapor can form at each point, that is, boiling can take place throughout the entire mass of water. Hence, at one moment the geyser is quiescent, and the next moment when the first bubble of vapor forms at the bottom, the pressure is relieved at every point of the liquid throughout the geyser, and every drop of water in the tube bursts into boiling.

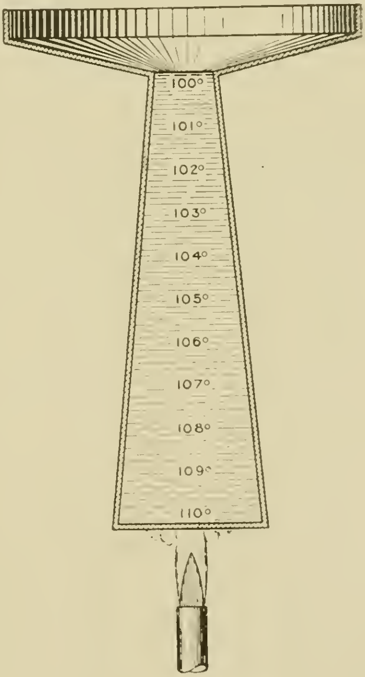

Fig. 173. The geyser.

The heat which raised the temperature of the water above 100° C. is now expended in producing a rapid evaporation of the water during the instant of eruption, and the water falls back again into the tube at the uniform temperature of 100° C. throughout. After an interval of time has elapsed sufficient to enable the water at each point again to be raised to the boiling point at that pressure, another eruption takes place, in which the above process is repeated, and the spouting of the geyser becomes perfectly periodic.

The action of the natural geyser, as well as the distribution of temperatures throughout its length, may be most advantageously studied by means of an artificial one, Fig. 173, consisting of a sheet iron tube filled with water, and a source of heat at the bottom.

265. The Spheroidal State. When drops of a liquid are placed upon a highly heated surface, as, for instance, when water is spilled upon a red-hot stove, it is observed that the liquid gathers itself into little globules which dart about over the heated surface without boiling. This phenomenon is due to the fact that the liquid is not in contact with the heated metal beneath, but rests upon a little cushion of its own vapor. That this is so may be shown by sending a ray of light parallel to the heated surface, when it will be seen that a small space separates the liquid drop from the metal, Fig. 174. Moreover, the rapid evaporation from the surface of the drop carries away heat from the liquid so rapidly that the temperature of the liquid forming the drop does not rise to 100° C.

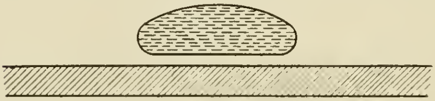

FIG. 174. A liquid in the spheroidal state.

When the liquid is in this condition, it is said to be in the *spheroidal state.* In the spheroidal state, therefore, a liquid is below its boiling point, and is not in contact with the containing vessel, but is separated from it by a delicate film, or cushion, of the vapor. This condition may be brought about only when the temperature of the vessel containing the liquid is very much above that at which the liquid boils. In consequence, therefore, of the extremely high temperature of the surface upon which the drop rests, and the exceedingly small distance that separates the two, evaporation takes place with enormous rapidity, particularly from the lower surface, and the volume of the liquid becomes rapidly less. So violent is the force with which the particles of vapor are projected from the lower surface of the globule, where it is nearest the heated surface, that the reaction of this force is sufficient to support the globule. The spheroid, therefore, hangs, as it were, in equilibrium between the downward force due to its weight and the upward force due to the reaction of the vapor particles as these are projected from its lower surface.

266. Solid Carbonic Acid Gas. Although under ordinary conditions carbonic acid is a gas, it may be liquefied by the application of very great pressure, and may be obtained commercially in this form in the heavy steel cylinders in which it is thus compressed. At ordinary room temperatures, the liquid exists in these cylinders under a pressure of about 60 atmospheres, which is its vapor tension at this temperature, and so, when the valve closing such a cylinder is opened, the liquid gas rushes out with great force.

In the air, the liquid stream encounters a pressure of only one atmosphere; and, as the gas could boil at any pressure less than 60 atmospheres, boiling and evaporation take place most energetically under this greatly reduced pressure. Inasmuch as no source of energy is present except the heat contained in the acid, the energy necessary to effect this rapid evaporation is taken from the heat contained in the liquid itself, and the stream of the liquid acid falls instantly and enormously in temperature.

If the cylinder of gas had been opened in a region where the pressure were 5 atmospheres instead of 1, the boiling and the evaporation, together with the consequent cooling, would have taken place until the temperature of the gas had fallen to —57° C. A condition of equilibrium would then have occurred in which the liquid would have remained in contact with its overlying vapor without further boiling. Moreover, if a quantity of the congealed gas at the same time temperature had then been introduced into the liquid, this solid gas would have remained unchanged and without effect in the liquid, that is, without melting and without causing a further solidification of the liquid gas.

Accordingly, at this temperature and pressure, the three states of this gas the solid, the liquid, and the gaseous, can exist in contact with one another without causing an increase of any one at the expense of the other two. The equilibrium condition thus determined by the pressure of 5 atmospheres and the temperature of —57° C., is known as the triple point for carbonic acid, and is the condition under which the boiling point and the freezing point of this gas fall together. If the pressure is increased above this amount, the solid will melt, while, if the pressure is reduced below this point, the liquid will boil at the expense of the heat which it contains; and as energy is thus withdrawn from this liquid at its freezing point, the gas will solidify.

Application of these principles may be found in the withdrawal of carbonic acid at atmospheric pressure from the steel cylinders above mentioned. If the liquefied gas at the temperature of the room is released from a cylinder where the pressure is 60 atmospheres, and collects in a vessel in the open air, where the pressure is 1 atmosphere, the gas is enormously above its triple point in temperature and below this point in pressure, hence rapid and violent boiling, evaporation, and cooling take place, until the liquid falls to the temperature of —57° C. This, however, is the temperature at which, under a pressure of 5 atmospheres, the boiling point and the freezing point of this gas fall together; but, as the gas is under a pressure of only 1 atmosphere, the process of boiling and further withdrawal of energy continues until the liquid is entirely frozen.

Even at this low temperature the now solidified gas is far removed from the absolute zero, hence its molecules possess still a large amount of kinetic energy. Therefore, at any pressure less than 5 atmospheres, the solid gas will evaporate, and will do so at the expense of the kinetic energy of its molecules. Its temperature thus falls; and, as evaporation continues, further energy is carried away from the solid gas, until, at atmospheric pressure, its temperature sinks to — 79° C.

This solid carbonic acid gas, which in many respects resembles snow, will remain for some time in the air, until the rapid evaporation which takes place from its surface finally reduces the substance to the gaseous state. Indeed, it is this rapid evaporation which is continually taking place that maintains the temperature of the solid gas at — 79° C. On account of the evaporation, it is impossible to raise its temperature at atmospheric pressure; for, if the attempt were made to heat it, the energy, which would then be added, would simply increase the evaporation, and the temperature of the solid gas would remain the same.

It is interesting to observe that this solid gas does not touch the table or other object upon which it may rest, since it is surrounded by a film of the vapor, produced by the rapid evaporation which is constantly taking place all about it. Although the temperature of this solid is — 79° C., it may nevertheless be handled with impunity, inasmuch as the hand does not come in contact with the solid, because it is prevented by this thin layer, or cushion, of vapor. If, however, this film of vapor were pressed out, so that the fingers were actually to touch the solid, the intense cold would then produce a frozen spot, resembling a burn.

It is partly because of this vaporous envelope that solid carbonic acid gas is a poor conductor of heat. For the same reason, objects surrounded by this white, snow-like material do not come into direct contact with the substance itself, and so are not readily cooled to its temperature. Alcohol or ether, when poured upon the gas, wet it, thereby displacing the vaporous layer around the particles, and forming a better contact between the gas and objects imbedded in it. As these liquids are soon brought to the temperature of the gas without forming a chemical union with it, objects which are immersed in this pasty mass quickly assume the temperature of the gas itself. In this way mercury, which solidifies at — 39° C., if placed in a metallic capsule and surrounded by this semi-fluid mixture, may be readily frozen, although the platinum crucible, in which the whole is contained, may be heated red-hot in the flame of a blast lamp.

267. Saturated Vapor. If the space above the surface of a liquid is enclosed, evaporation will take place in the manner described (p. 275), until the molecules gather in great numbers in this closed region. These particles of vapor behave in every respect like the molecules of a gas, and collide with one another and strike against the walls of the vessel, thereby causing a pressure. Because of these molecular collisions, it often happens that one particle, in colliding with another, projects the latter back again into the liquid from which it originally came. *When just as many particles thus return to the liquid as leave it, a condition of equilibrium is reached, and the vapor is said to be saturated.* A saturated vapor can be formed only in a closed space above a liquid, because, if this region were open, diffusion and currents of air would rapidly carry away the vapor. A closed space above a liquid surface, therefore, is always filled with the saturated vapor of the liquid at that temperature.

As long as the vapor in this space is not saturated, the vapor pressure, or tension, increases as evaporation continues. When, however, the vapor is saturated, *this saturated vapor exerts a perfectly definite pressure, which is known as the pressure, or tension, of the saturated vapor for that particular temperature.* If the temperature of the liquid is raised, more of the liquid will evaporate into the closed space, the kinetic energy of the molecules will likewise increase, and the pressure of the vapor will increase. The pressure of the saturated vapor, accordingly, depends wholly upon the nature of the liquid and upon the temperature to which it is raised, and not at all upon the extent of the space which is filled with the vapor. The pressure of water vapor at 100° C., is exactly one atmosphere, because, when water boils, the particles of the water vapor must push back the atmosphere, and hence must exert a pressure just equal to the atmospheric pressure. In general, *the maximum pressure of a saturated vapor at any temperature is the pressure at which the liquid will boil at that temperature.*

268. The Pressure Exerted by a Saturated Vapor. The fact that the pressure produced by the saturated vapor of one liquid is different from that produced by the saturated vapor of another liquid at the same temperature may be illustrated by providing three barometer tubes, Fig. 175, and introducing into the first a few drops of water, into the second a few drops of alcohol, and into the third a few drops of ether. Each of these liquids rises to the top of its mercury column, and there evaporates very rapidly into the Torricellian vacuum. If, then, a small drop of the liquid remains above

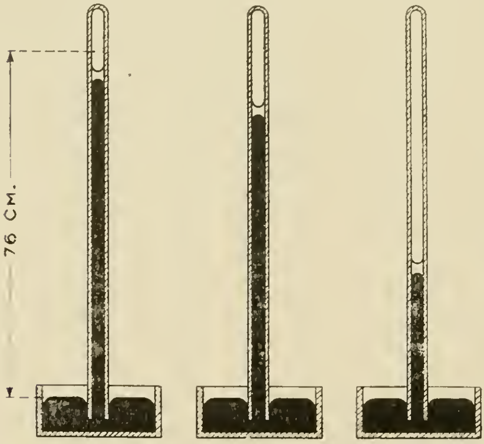

FIG. 175. Pressures exerted by saturated vapors.

the mercury column, the space above is filled with the saturated vapor of that liquid. The column containing the water falls slightly in height, showing that at ordinary temperatures the water vapor above the mercury column exerts a relatively small, though noticeable, pressure, the amount of this fall indicating the pressure of saturated water vapor at that temperature. The depression of the column in the tube containing the alcohol is greater than that in the tube containing the water, and the depression in the tube containing the ether is very much greater than in that containing the alcohol. Indeed, the pressure of saturated ether vapor at the ordinary room temperature is nearly two-thirds of an atmosphere.

If, now, these three tubes are slightly warmed, the depression in each case is seen to increase, showing that at a higher temperature more of the liquid evaporates into the space above the surface, and that the pressure of a saturated vapor at a high temperature is greater than that of the same saturated vapor at a lower temperature.

In this connection, it must be borne in mind that *if a saturated vapor is cooled, some of this vapor will be condensed to the liquid state*, and the uncondensed portion will remain saturated at the lower temperature, and exert a diminished pressure.

269. Dalton's Laws. Dalton found that if the closed space above a liquid is filled with air or other gas that does not react chemically, just as much liquid will evaporate into the space as would evaporate if this region were a vacuum, except that longer time will be required. That is, every vapor will exert its own pressure independently, although other gases or vapors may be present. If, therefore, any closed space is filled with a mixture of several vapors that do not react upon one another, the resulting pressure due to all these vapors is equal to the sum of the pressures that each vapor would exert if it occupied the same space alone. These two statements are known as Dalton's laws.

270. The Pressure of a Saturated Vapor is the Maximum Pressure. If an unsaturated vapor is compressed, an increase of pressure will produce a decrease of volume. An unsaturated vapor, therefore, behaves in every respect like a gas, and within narrow limits obeys Boyle's law. If, however, the attempt is made to apply pressure to a saturated vapor, it is found that, as the volume becomes less, the pressure does not increase, but that the vapor pressure and the vapor density remain perfectly constant, and that the diminution of volume simply forces some of the saturated vapor over into the liquid state. Since it is impossible to increase the pressure upon a saturated vapor, it is evident that a saturated vapor does not obey Boyle's law.

This may be shown by passing a few drops of ether into a barometer tube which is inverted over a deep cistern of mercury. The pressure of the saturated ether vapor within the tube will depress the column of mercury about 46 centimeters, Fig. 176, leaving the surface of the mercury within the tube only about 30 centimeters above the surface of the mercury in the cistern. If, now, the tube is lowered in the cistern, the volume of the vapor will be decreased, but its pressure, as indicated by the unchanging height of the column of mercury, will remain perfectly constant until the vapor has all been condensed to the liquid form. The pressure upon a saturated vapor, therefore, cannot be increased, but always remains constant, and is independent of the volume and of any variation which the volume may undergo. Moreover, *the pressure which a saturated vapor exerts is always the maximum pressure which the vapor can exert at that temperature.*

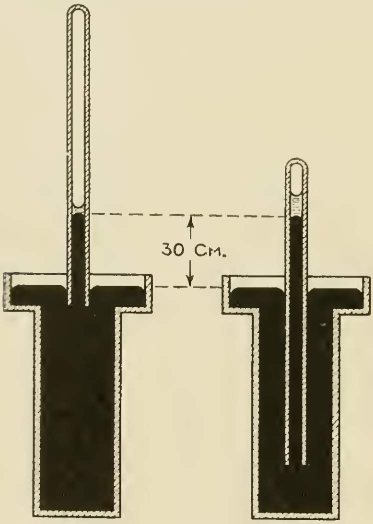

FIG. 176. A saturated vapor exerts its maximum pressure.

271. Dust Particles or Other Nuclei are Necessary for Condensation. If a saturated vapor is quickly cooled, as, for instance, if it is allowed suddenly to expand (p. 303), a portion of this vapor is condensed to the liquid form, and appears as a cloud, or fog, which slowly settles to the bottom of the containing vessel. This precipitation takes place more readily, however, when the air in which the vapor is suspended is in its natural condition, that is, while it still contains the multitude of dust particles with which it is usually filled. On the other hand, if this air is first freed from all traces of dust and other substances by being drawn through a tube filled with loosely packed cotton, no condensation takes place when cooling occurs, and no cloud is formed. The vapor is then said to be supersaturated, for a given volume, even at the lower temperature, contains just as great a quantity of vapor as was sufficient to saturate this volume at the higher temperature.

Experiments like these point unquestionably to the interesting fact that the condensation of a vapor, that is; the passage from the state of saturated vapor to that of a liquid, is facilitated by the presence in the vapor of minute particles of dust or other foreign matter, upon the surface of which condensation begins. If these particles, or nuclei of condensation, are present, a portion of the vapor will be precipitated as soon as the temperature falls and the remaining vapor will then be saturated at the lower temperature. If, however, these nuclei are absent, as is the case in dust-free air, condensation is produced less readily upon a lowering of the temperature. In this case the vapor, which is then supersaturated, possesses a greater density and exerts a greater pressure than would be possible at this temperature if condensation could occur.

Moreover, since small particles ordinarily exist in the vapor before condensation begins, it follows that the number of droplets produced depends in large measure upon the relative number of these nuclei that may be present. The frequent occurrence of heavy fogs in London and other densely populated cities is thus seen to be due to an excessive contamination of the air by dust and smoke, each particle of which serves as a center for the condensation of the saturated water vapor of the atmosphere.

272. The Critical Temperature. When a liquid evaporates into a closed space, volatilization takes place, as we have seen (p. 287), until the closed region above the liquid surface is filled with the saturated vapor. If the temperature is raised, more of the liquid evaporates, and the vapor, which is then saturated at a higher temperature, becomes denser, and exerts a higher pressure.

The rise of temperature, however, which thus produces an increased density of the vapor, occasions likewise, a decreased density of the liquid, inasmuch as the liquid expands and so occupies a larger volume. With a further rise of temperature, this double process continues, the density of the vapor becoming constantly greater and that of the liquid continually less, until, for each liquid, a perfectly definite temperature is reached, at which the densities of the vapor and the liquid become equal. At this particular temperature, which is called *the critical temperature,* the properties of the vapor and the liquid become in every way identical. Moreover, the bounding surface of the liquid, which, during the heating, has been marked by a clearly-defined meniscus, suddenly vanishes, the distinction between the liquid and the vapor completely disappears, and the tube in which this process has been conducted is uniformly filled.

At any temperature below the critical temperature for a particular liquid, a decrease in the volume of the saturated vapor will produce no increase of pressure, but will result in condensing a portion of the vapor, that is, in reducing it to the liquid form. At any temperature above the critical temperature, however, a decrease of volume will give rise to a corresponding increase of pressure, and this increase of pressure may be carried to any desired extent without producing liquefaction. It is, therefore, impossible to liquefy a gas at a temperature above its critical temperature. Indeed, it is not until a gas has been lowered in temperature to a point below its critical temperature that it is possible to liquefy this gas by pressure. The critical temperature is therefore sometimes used to indicate the distinction between a gas and a vapor, a gas being a vapor above its critical temperature, while a vapor is a gas below its critical temperature. So fundamental and definite is this characteristic, that *the critical temperature is defined as that temperature above which no pressure, however great, will reduce a gas to the liquid state.*

CHAPTER XX.

THE HUMIDITY OF THE ATMOSPHERE.

273. Dew. Owing to the evaporation which is continually taking place from all exposed bodies of water and from the moistened surface of the earth, the atmosphere always contains water vapor as one of its constituents. The amount of this moisture, which varies greatly at different times, depends upon the temperature, upon the proximity of large bodies of water, and upon winds which carry the vapor to regions remote from the place where evaporation occurs.

Under the action of bright sunshine during the morning hours following a rainy night, evaporation from the earth takes place with great rapidity; and as the temperature rises, the amount of vapor which thus passes into the air is increasingly great. Although the air contains large quantities of water vapor during the middle portions of the day, this amount is, in general, wholly insufficient to produce saturation, owing to the high temperature which then prevails. As the air becomes cooler towards night, however, the temperature is approached at which the vapor present will become saturated, and this vapor is nearing its point of condensation. The temperature at which this occurs is known as the dew-point. Accordingly, if the air is cooled below its dew-point, some of the moisture will be condensed and precipitated, and the air will then remain saturated at a lower temperature. This explains the occurrence of the fog which is sometimes seen over flat meadows on quiet summer evenings.

The air, however, is a poor conductor of heat, hence it more usually remains at or near its point of saturation, until it is cooled locally by coming in contact with a much colder object, upon which the moisture is then deposited. Blades of grass and leaves of foliage are good radiators and relatively poor conductors of heat, so that during the night, if the sky is clear, they radiate into space the heat they absorbed during the day, and so become many degrees colder than the surrounding air. The saturated air, as it comes in contact with these cold objects, is itself cooled, and the moisture which it contained at the higher temperature is deposited in the form of minute particles, which then coalesce to form a drop of dew. It is probable also that the amount of dew is materially increased by the condensation of the moisture which evaporates from the still warm earth, as this meets the cold foliage immediately above it.

If, however, the night is not clear, the clouds prevent the radiation out into space which would otherwise take place, hence objects near the surface of the earth do not fall so low in temperature, and therefore cannot condense the moisture of the surrounding air. In other words, the clouds radiate back to the earth nearly as much heat as the earth radiates to them, hence there is but a small fall of temperature, and objects do not become sufficiently cold to condense the moisture of the air. It is on this account that but little dew "falls" on a cloudy night. Occasionally low, flat fields of delicate plants may be protected from frost on clear, windless nights by building smoky fires around the field. The smoke partially covers the field and so, to a certain extent, acts like the clouds in preventing so great a radiation from the plants as to cause them to fall in temperature possibly below the freezing point. The production of hoar frost is similar to that of dew, except that it occurs at temperatures below the freezing point of water.

274. Humidity. The state of the air as regards the amount of moisture it contains is known as its humidity. If the quantity of water vapor present is so great that it is near its point of saturation, the humidity is said to be high, but if this moisture is relatively small in amount, the humidity is low. More accurately, however, the word humidity is used in two senses, distinguished as absolute and relative. By the absolute humidity is meant the actual number of grams of water that are present at any time in a cubic meter of air. By the relative humidity is meant the ratio of the number of grams of water that are actually present in a cubic meter of air to the number of grams that would saturate this quantity of air at the same temperature. Inasmuch, however, as the pressure, or tension, which an unsaturated vapor can exert is proportional to the amount of the vapor present, *the absolute humidity* may be expressed as the pressure in millimeters of mercury which the unsaturated water vapor present exerts, and *the relative humidity* is then the ratio of the tension of the water vapor, as it is found to exist in the air, to the tension which saturated water vapor would exert at the same temperature.

275. Hygrometry. The humidity of the air, both absolute and relative, may be measured in a number of ways, and the science which thus deals with the determination of the humidity is known as hygrometry. The absolute humidity may be measured directly by weighing tubes containing substances which readily absorb moisture, such as sulphuric acid or calcium chloride, and then drawing a measured quantity of air through these tubes. The increase of weight which thus results is the amount of water vapor that was contained in the measured quantity of air, and from this the absolute humidity may be calculated. From tables it is possible to learn the amount of moisture that would saturate the air at any temperature, hence the relative humidity may be readily determined when the absolute humidity is known.

A second method of measuring the humidity consist in determining the dew-point of the air, and then computing the humidity from tables which have been prepared for this purpose. The dew-point of the air, that is, the temperature at which the water vapor in the air becomes saturated, may be approximately determined by filling a bright tin cup with water at the temperature of the air, and gradually pouring in ice water from a pitcher until a slight clouding occurs on the bright surface of the cup. The temperature at which this takes place is the temperature at which the water vapor in the air is precipitated upon the cold surface, that is, the temperature at which dew forms. In other words, this is the temperature at which the water vapor present in the air becomes saturated. From the known temperature at which the dew-point occurs, the humidity of the air may be calculated.

Another method, more accurate than the above, consists in filling with ether a metal box which is polished and gilded on the outside, and drawing air through this box, thus causing the ether to evaporate. The ether cools as it evaporates, and the metal box falls in temperature until the moisture of the air is deposited upon it, and a clouding of the bright surface occurs. A thermometer placed in the ether gives the temperature at which this takes place. The current of air passing through the ether is then stopped and the box allowed to warm from the air, until the cloud of dew upon the polished surface vanishes. The temperature is again read and the mean of the two gives most accurately the temperature of the dew-point.

276. The Wet and Dry Bulb Hygrometer. The most convenient and the most usual form of hygrometer is the one devised by August, and known as August's psychrometer. This consists of two exactly similar thermometers, one of which is exposed freely to the air, and which therefore gives the temperature of the air, while the bulb of the other is surrounded by a muslin gauze kept wet by means of a wick that dips into a cup of water beneath. Evaporation, therefore, takes place from the moist surface of the muslin, and the heat necessary to produce this evaporation is withdrawn from the bulb of the thermometer, thereby lowering its temperature.

The amount of evaporation, and therefore the amount of cooling of this thermometer, depends upon the amount of water vapor present in the air. If the air is saturated, it can contain no more moisture at that temperature, no evaporation will then take place, and the wet bulb thermometer will indicate the same temperature as the one with the dry bulb; but if the air is far removed from its point of saturation, evaporation, and therefore, also, cooling will take place with great rapidity from the wet bulb, and this thermometer will indicate many degrees lower than the other. From the known temperature of the air, and from this observed difference in the readings of the two thermometers, the dew-point at that temperature and the absolute as well as the relative humidity may be computed with the aid of tables which have been prepared.

277. Winds. The unequal and the variable distribution of heat over the surface of the earth is the fundamental cause of the movements of the masses of air taking place continually in all quarters of the globe. If the air in any particular locality is heated, it expands, and as it thus becomes less dense, it rises, while cooler, denser air from the sides flows in along the surface of the earth to take its place. Convection currents on an enormous scale, therefore, constantly tend to equalize the temperature of the earth, the heat of the tropics being thus carried to the cooler latitudes, and the cold air from the polar regions being brought down to moderate the heat of the torrid zone. The passage of these masses of air near the surface of the earth produces the phenomenon of winds.

While many local disturbances intervene to change the direction of the winds and to produce anomalous movements of the air, the winds in the northern hemisphere may thus, in general, be divided into two great classes, namely, those which blow north and those which blow south, for the reason just given. The air, however, at any part of the earth acquires the velocity of the earth at that place, and moves with the earth in its diurnal revolution. In the topics, therefore, the heated air that rises and flows northward carries with it the high velocity of the earth in the equatorial region, so that as it moves northward it comes to parts of the earth that have a smaller eastward velocity. The air, accordingly, runs ahead of the earth, so to speak, that is, its high eastward velocity, combined with its motion northward gives it a movement towards the northeast, or, in other words, this wind blows from the southwest. On the other hand, the cold air from the higher latitudes, as it flows southward to take the place of the air which has risen, carries with it the low velocity of the polar regions from which it came, so that it continually comes to parts of the earth that have a higher eastward velocity. In this case, therefore, to continue the same figure, the air constantly lags behind the earth, and the movement southward compounded with this slower motion eastward gives this polar air a relative motion toward the southwest, that is, the wind blows from the northeast. This warm air blowing from the southwest and the cold air from the northeast are the main causes for the occurrence of storms, as we shall see later.

278. Clouds. If the heated air, as it rises, happens to be over or near large bodies of water, evaporation takes place with great rapidity and large quantities of water vapor are absorbed by the air. Notwithstanding the high temperature and consequently the very high point of saturation, the air over the tropical oceans is very nearly saturated with water vapor. As this heated, moist air expands and rises, the vapor which it contains is likewise carried into cooler regions, so that moisture also, as well as heat, is transferred as a consequence of this convection. When this warm, vapor-laden air falls in temperature below a point at which the vapor present will be saturated, a part of the moisture is condensed, thus forming a cloud, and the air remains saturated at this lower temperature.

The causes which lead to the cooling of this mass of warm air and the consequent condensation of a portion of its moisture, are two-fold, and must be carefully distinguished. A cooling will take place when this mass of warm air meets colder masses of air, or comes in contact with cold currents from the north, and a condensation must then necessarily follow. This effect, however, takes place principally around the edges of the mass of hot air, so that the interior portions are comparatively unaffected in temperature.

A much more important cause, however, and one which is far more effective in producing a lowering of the temperature and a condensation of moisture, is the fact that this warm air expands as it rises and in expanding it has to do work in pushing back the atmosphere that lies all around it. Inasmuch as no other source of heat is present, the energy in the form of heat which is necessary thus to do this external work is taken from the heat contained in the warm air itself, and accordingly the mass of air becomes cool. The air, therefore, is cooled by its own expansion, that is, by doing work in pushing back the atmosphere to make room for the increased volume which the expanded air must occupy. Indeed, the cooling by this process is so great that the air frequently falls in temperature below the dew point, and the vapor present is partly condensed into a multitude of minute particles forming a cloud.

The formation of a cloud may be observed almost any summer morning when a bright, clear day follows a rainy night. The warmth of the sun causes rapid evaporation from the moist earth, and columns of heated air, laden with water vapor almost to the point of saturation, rise from some spot, such as the side of a hill, where the formation of an upward current is particularly favorable. This heated air expands as it rises, and cools as it expands, until it falls below the temperature at which the vapor present in the air becomes saturated, and then a cloud occurs. The large, fleecy, cumulus clouds which appear in the sky by nine or ten o'clock in the morning are simply the tops of these rising columns of air which have been cooled by expansion, until a portion of the vapor which the air contains has thus been condensed.

Contrary to the popular belief, these particles of moisture which have been condensed from the vapor are little spherical globules, about the hundredth part of a millimeter in diameter, and are not hollow, as is generally supposed. They are so exceedingly small, and their friction with the air is so great that they fall only very slowly, and the slightest current of air is sufficient to carry them upward. Indeed, if the particles at the lower edge of the cloud were to fall, they would soon come into a warmer region where the point of saturation is higher. They would then immediately evaporate and thus form a part of the vapor in the rising column of air, only again to be condensed as soon as they came once more to the height from which they fell.

If the upward current is very strong, as may be the case over a highly heated locality, the rising column of air may pass far above the point where condensation begins, and soon may be cooled by its expansion to a temperature so low that the vapor freezes as it condenses. Small crystals are then carried upward by the still rising column of air, and at a height of five or six miles, when the upward current has ceased, these crystals may be seen forming the thin, filmy masses known as cirrus clouds, which are the highest clouds known. The light haze which often overspreads the sky is due to these masses of ice crystals at this great height, forming a thin diffused cirrus cloud.

After clouds are formed a still further cooling produces an increased condensation, until the number of these tiny particles becomes so great that they collide and coalesce and form a drop too heavy to be supported by the currents of air, and this then falls as rain. Snow occurs instead of rain when the tiny water particles come together and unite at a temperature below the freezing point of water.

279. The Weather. The movements of these heated, moisture-laden masses of air from the south, and of the cold, dry currents from the north, modified, of course, by many local conditions, are the main causes for the formation of the storms which sweep across the country with greater or less violence. The heated air rises because of the smaller density in its expanded condition, and consequently, in regions where these upward currents take place, the pressure of the air is less than it otherwise would be, and the barometer here stands very low. The heated column of moist air, and with it also this area of low barometer, now moves forward from the southwest, while the cold, dry air from the polar regions moves down from the northeast, and with it also comes a region of higher barometric pressure.

The meeting of these two great masses of air from opposite directions gives rise to a rotary motion, forming a great vortex, often several hundred miles in cross-section, and this moves in a general northeasterly direction, requiring perhaps three or four days to travel across the country. This great vortex is known as a cyclone, without, however, carrying with it the idea of the destructive storm usually called by this name, which is more properly a tornado. The direction in which this cyclone revolves, as viewed from above, is opposite the hands of a watch, and the region of lowest barometric pressure is always at its center. This cyclone is therefore usually spoken of as a "low," and appears as such on the daily weather maps. On the other hand, the cold, dry air from the north acquires also a rotary motion, but in the opposite direction, that is, in the direction of the hands of a watch, and in the center of this region is the area of highest barometric pressure, so that this area, or anti-cyclone, as it is called, is more usually expressed as a "high."

Simultaneous observations of the atmospheric conditions are taken daily all over the country; and when the barometric height is plotted on a map, the points at which this height is the same may be connected by a line, known as an isobar. The isobars, which are closed curves about a low, indicate the position of the center of this low, and enable its motion forward to be studied. The change in the barometric height for a given distance on the surface of the earth is known as the gradient. The nearer together the isobars occur, the steeper is the gradient, and the greater the velocity of the wind.

280. Weather Prediction. Because the great storms which sweep across the country are produced in the majority of cases under similar circumstances, and because these storms have a general easterly, or northeasterly movement, the changes in the weather which are thereby produced are characteristic and quite definite. The air over a heated region in the southern and southwestern states becomes itself heated, expands, and rises, while the cooler air flows in from the sides to take its place. A cyclonic area is thus produced, the direction of rotation of this cyclone, as viewed from above, being opposite the hands of a watch, and the center of this area being the region of lowest barometric pressure.

As the air expands in the rising column at the center, it must do external work in pushing back the atmosphere, and the heat thereby consumed is withdrawn from the air, so that its temperature falls. As soon as the dew point is reached, a cloud is produced; and if, as is generally the case, the central upward current is very strong, the ascending column may rise to the height of five or six miles, when the moisture will be congealed into crystals of ice forming a cirrus cloud. These cirrus clouds are then carried to the northeast by the winds from the equatorial regions, so that in this way the approach of a storm may be heralded two or three days in advance of its arrival.

At a point northeast of the storm center, the winds then begin to blow with increasing violence from the southeast and south, and the air becomes hot and sultry. The cirrus clouds soon spread into a general cirrus haze, and this in turn yields to the heavy, dense clouds which are formed by the condensation of the vapor in the moisture-laden winds, as they blow from the southern states and from the Gulf of Mexico, where the humidity is very high. As this moisture is still further condensed, rain begins to fall and increases in quantity as the region of lowest barometric pressure approaches. When, however, the storm center passes, the wind begins to blow from the northwest, cold, dry air is thus brought down from the north, the clouds soon dissolve and disappear, bright sunshine, with cool, refreshing weather is brought about in summer and a cold wave prevails in winter. The condition continues for a few days, when these processes are repeated in much the same order.

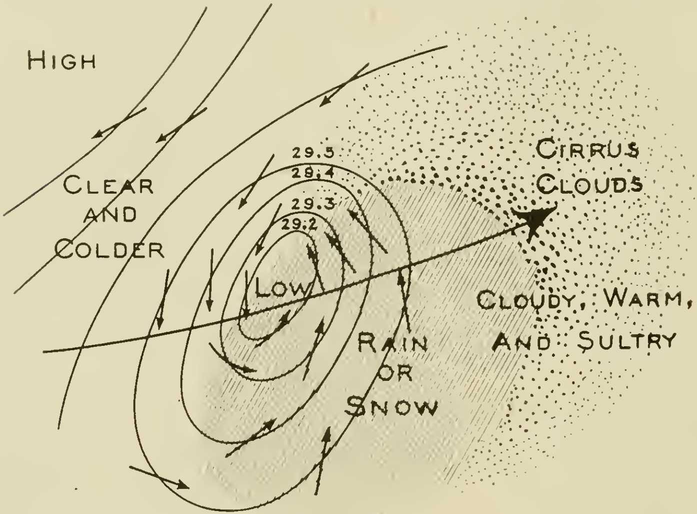

FIG. 177. The conditions existing about a storm center from which the weather predictions are made.

If a storm center passes south of a particular region, the winds at that point will change from east through the northeast and north to northwest, and the warm, cloudy weather will slowly yield to the dry, cold, clearing winds from the northwest. If, however, the storm center passes far to the north, the winds will change from the east through the southeast and south to the southwest, and then later to the west, and a succession of warm, pleasant, sunny days is the result.

The movements of the air, as well as the atmospheric conditions prevailing over the country at any time may be most advantageously studied by means of the weather maps which are issued daily, and which set forth the meteorological conditions then existing, Fig. 177. From these conditions, and from a knowledge of the general movement which the storm center will probably follow, the weather indications for any particular locality may be predicted with a fair degree of certainty, often for twenty-four hours and even longer in advance.

CHAPTER XXI.

THERMODYNAMICS.

281. Heating by Compression and Cooling by Expansion. When a gas is compressed by an external force, as, for instance, when a piston is pushed down in a cylinder, an amount of work must be done equal to the product of the average force thus exerted and the distance through which the piston moves. This amount of energy is added to the energy of the gas, the kinetic energy of the molecules is thereby increased, and the gas becomes heated. When, however, a gas expands, and in so doing performs work, that is, for example, when the gas is allowed to push the piston up against an outside force, the energy which is thus consumed is taken from the kinetic energy of the molecules, and the gas becomes cool. A gas, therefore, is heated when it is compressed by an outside force, and cooled when it expands against an outside force, since work is done upon the gas in the first instance and by the gas in the second.

The equivalent of pushing up the piston is accomplished by suddenly opening the valve in a reservoir of compressed gas, and allowing the gas to expand into the atmosphere or into another reservoir containing gas at a lower pressure. In this case the cooling effect is a complicated one, and, in truth, consists of three effects combined, all of which withdraw energy from the gas, and all of which, therefore, aid in producing the observed lowering of temperature.

In the first place, the gas in expanding into the air must push back the atmosphere, thereby performing an amount of work which may be called external work. Since no other source of energy is at hand, this work is done at the expense of the kinetic energy of the molecules, and the temperature, therefore, is lowered.

A second cooling effect is produced when expansion occurs, due to the fact that a given particle of the compressed gas pushes out another particle of the gas immediately before it, that is, it imparts a velocity to this particle, and therefore does work. Moreover, each particle, in turn, as it approaches the opening in the reservoir, yields a portion of its energy thus in communicating a higher velocity to the particle before it, so that the energy of the gas remaining in the reservoir is diminished, and this gas grows cold. The increased velocity which the particle thus acquires in leaving the reservoir is a directed velocity, and produces a jet, or wind, of the escaping gas. This must be carefully distinguished from the wholly irregular motion of the particles due to molecular bombardments, for the directed linear velocity of the particles constituting the wind has no effect upon the temperature.

A third cause which brings about a cooling upon expansion is the fact that no known gas is a perfect gas, and that at ordinary temperatures the particles of all gases have a very slight attraction for one another. When expansion occurs, therefore, work must be done in separating the particles of gas from one another; that is, internal work must be performed in changing a portion of the molecular kinetic energy of the gas into molecular potential energy. For air at room temperatures, the cooling which results from this effect amounts to about $\frac{1}{4}^\circ$ C., for each atmosphere of expansion, and is known as the Joule-Thomson effect, in honor of its two noted discoverers.

282. The Liquefaction of the Air. Air, when compressed to about 200 atmospheres and allowed suddenly to expand, is therefore considerably cooled by the withdrawal of a portion of its internal energy. If this cooled air is allowed to circulate about the tube containing the compressed air which is about to expand, the latter will be cooled almost to the temperature of the expanded air before its own expansion occurs. When this in turn, expands and cools still further, it may likewise serve to surround and cool still other air which is about to expand; and when this process is continued, so intense a degree of cold is finally produced that the air in expanding falls to the temperature of —191° C.

This is the temperature at which the saturated vapor of air exerts a pressure of 1 atmosphere, or, in other words, it is the average temperature at which the gases of the air can boil at atmospheric pressure, the temperature of boiling oxygen being —182.5° C., and that of boiling nitrogen —198° C. Accordingly, these gases of the air, as they expand and cool, condense to the liquid form, and collect as boiling air in a vessel placed to receive them. The operation here described, which is called the regenerative process, is the practical method of liquefying the air at the present time, and is a common means by which liquid air may be produced in large quantities.

The inevitable access of heat, which even the use of the Dewar flask (p. 251) cannot altogether avoid, causes this exceedingly cold mixture continually to evaporate, if not actually to boil; and, as the nitrogen boils at the lower temperature, it volatilizes the more rapidly and disappears the more quickly, although the oxygen evaporates at a rate but little slower. Accordingly, as a quantity of liquid air is tested from time to time, it will be found to contain a continuously decreasing amount of nitrogen, that is, it will become constantly richer in oxygen, until, after two or three days, the remaining liquid will be found to be very nearly pure liquid oxygen.

When hydrogen at ordinary temperatures is allowed to expand, it is heated instead of being cooled. But if it has been previously cooled to —80.5° C. and then allowed to expand, a cooling begins, which is still further increased by the regenerative process. The gas then falls in temperature to —253° C., which is its boiling point at atmospheric pressure, and therefore liquefies as it expands. If liquid hydrogen is evaporated at reduced pressure, its temperature falls still lower, until it solidifies at —257° C. Helium gas, when expanded by the regenerative process, is lowered in temperature until it liquefies at —268.7° C.; and when this liquid is evaporated at a pressure of 1 centimeter of mercury its temperature sinks to —270° C., only 3° above the absolute zero.

283. The Isothermal Compression of a Gas. If we consider a mass of gas enclosed in a cylinder provided with a piston, we may diminish its volume by applying pressure to the piston; and if the gas is kept at constant temperature, the compression takes place according to Boyle's law. We have seen, however, that a mass of gas is warmed by compression and cooled by expansion. Therefore, if the gas is to remain at a constant temperature, we must assume that the walls of the cylinder are of a material which is a good conductor of heat, so that the heat which is produced by the compression is conducted away through the walls as fast as it is developed, and the enclosed gas remains at its initial temperature, however its volume is made to vary. Under this condition of perfect constancy of temperature, a change in the pressure and the volume of a gas is known as an *isothermal* change.

The condition of a mass of gas at all times as regards its pressure and volume becomes at once evident, if we construct a diagram consisting of two straight lines at right angles to each other, Fig. 178, the horizontal one, for instance, representing the axis of volumes, and the vertical one representing the axis of pressures. Distances then laid off from the point of intersection of these lines represent volumes and pressures respectively, one division on the horizontal line representing one liter, two divisions, two liters, etc., and one division on the vertical line representing one atmosphere, two divisions two atmospheres, etc. With the aid of a diagram such as this, the entire condition of a mass of gas as regards its volume and pressure at any given temperature may be represented by a single point, because perpendiculars dropped from this point upon these axes indicate the volume and the pressure respectively. That is, if the perpendicular Pv is dropped from the point P upon the horizontal axis, which is the axis of volumes, the intercept Ov upon the axis represents the volume of the gas. In the same way, if the perpendicular Pp is dropped from the point P upon the vertical axis, which is the axis of pressures, the intercept $O\,p$ represents the pressure of the gas.

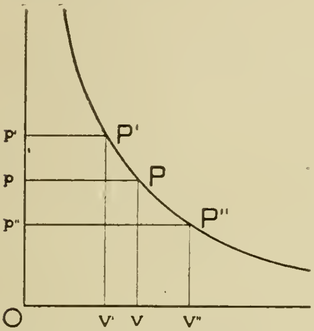

FIG. 178. The isothermal compression of a gas.

If now the piston is pushed down thus compressing the gas in the cylinder, the volume is diminished while the pressure at the same time is increased, and the condition of the gas is now represented by the point P', where Ov' represents the new volume and $O\,p'$ the new pressure. But this compression has been performed at constant temperature, therefore Boyle's law has been obeyed, that is, the product of the pressure and the volume must be constant. Accordingly the area of the rectangle $OvPp$ just equals the area of the rectangle $Ov'P'p'$, or,

$$Op \times Ov = Op' \times Ov'.$$

If the piston is drawn up, so that the gas occupies a larger volume, the pressure is diminished, and the condition of the gas as regards its volume and pressure is indicated by the point P'', where Op'' represents the pressure of the gas when the volume has become Ov''. Here again

$$Op \times Ov = Op'' \times Ov''.$$

284. Isothermal Lines. During the above changes the point P has moved along the curve $P'\ P''$, which is of such nature that at each point it represents Boyle's law, or

$$P\,V = \text{constant}.$$

This expression may then be called *the equation of the curve,* where P and V are the co-ordinates, since at each point it expresses the condition which the curve represents. There is, however, but one curve which has an equation of this form, and that is the rectangular hyperbola. It follows, therefore, that, as compression and expansion take place according to Boyle's law, the point P, which indicates the condition of the gas at that temperature, moves along a rectangular hyperbola, of which the axes of pressure and volume are the asymptotes. An asymptote to a curve means a line such that a point which moves along the curve continually approaches that line but never meets it. In the compression of the gas this indicates that however large the volume may be, there will always be some pressure exerted by the gas, for whatever distance P moves to the right, it will still be appreciably above the axis of volumes. Again, if the pressure on the gas is increased, the volume will diminish, but, to whatever extent P may move upward and to the left, it can never meet the axis of pressures, for the mass of gas always has some volume, however great the pressure to which it is subjected. A curve such as $P'P''$, every point of which represents the instantaneous condition of the gas at constant temperature, is known as *an isothermal curve, or an isothermal.*

If this process of change of volume and pressure is carried on at a higher temperature, then for a given volume the pressure will be greater, and another isothermal curve will be traced, Fig. 179, lying above the first and similar to it, but nowhere intersecting it. In this way a whole family of isothermal curves may be drawn, a separate curve being necessary for each temperature at which the compression or expansion is conducted.

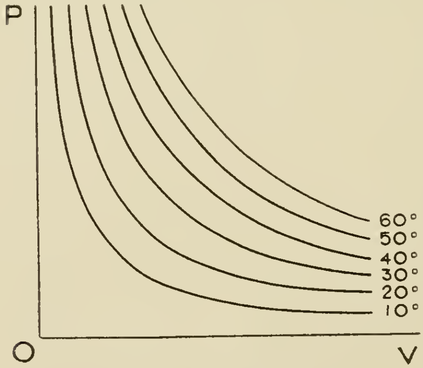

FIG. 179. Isothermal lines.

The very great convenience of these isothermals lies in the fact that, of the three variables, pressure, volume, and temperature, herein employed, if two are given, the third may always be determined.

1. If the pressure and the volume of a gas are given, the temperature may be found in the following way: At the given pressure erect a perpendicular to the axis of pressures, and at the given volume erect likewise a perpendicular to the axis of volumes. The intersection of these two lines will lie on the isothermal denoting the temperature at which the gas having that particular volume will also have that particular pressure.

2. If the pressure and the temperature of a gas are given, the corresponding volume may be determined by the following method: At the given pressure erect a perpendicular to the axis of pressures, and extend this line until it intersects the isothermal of the given temperature. A perpendicular let fall from this point of intersection upon the axis of volumes will intercept upon this axis a distance from the origin which will be the volume required.

3. If the volume and the temperature of a gas are given, the corresponding pressure may be found as follows: At the given volume erect a perpendicular to the axis of volumes and extend this line until it intersects the isothermal of the given temperature. A perpendicular let fall from this point of intersection upon the axis of pressures will intercept upon this axis a distance from the origin which will be the pressure required.

285. The Elasticity of a Gas. If we have a series of isothermals for gas, a single point such as *P* will not only indicate the pressure and the volume of the gas for that temperature, but will also enable us to determine the elasticity of the gas under these same conditions of temperature, pressure, and volume. For, if we again imagine a mass of gas enclosed in a cylinder provided with a piston, the point *P* on the isothermal curve, Fig. 180, will represent the condition of the gas as to volume and pressure at that temperature, because the perpendiculars let fall from *P* on the two axes show that the gas has a volume represented by Ov and a pressure represented by Op.

If now, the piston is pushed down, the volume of the gas will be diminished to some new volume such as Ov', and the pressure will be increased to some new value such as Op'. During this change the point P will move to P' along the line PP', which is a portion of the isothermal for that temperature, and which is so short that it may for the present be considered as straight. The perpendicular Pp intersects the perpendicular $P'v'$ in M, so that PM, which equals vv', denotes the change in the volume of the gas, and $P'M$, which equals $p'p$, represents the change in the pressure.

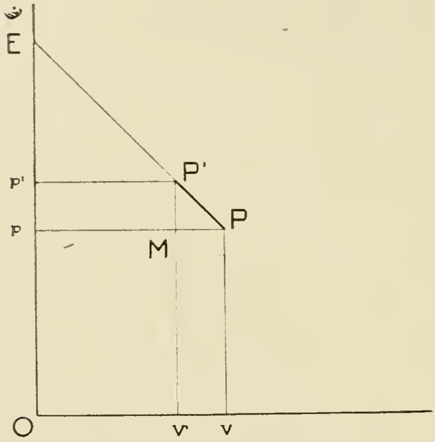

Fig. 180. The elasticity of a gas.

We now define the voluminal compression of a gas to be the ratio of the change of volume to the original volume. That is, on the diagram, PM represents the change of volume of the gas, and pP, which is the same as Ov, represents the original volume. Therefore,

$$\text{Voluminal compression} = \frac{PM}{pP}.$$

We also define the elasticity of a gas to be the ratio of the increment of pressure to the voluminal compression produced. That is,

$$\text{Elasticity} = \frac{\text{Increase of pressure}}{\text{Voluminal compression produced}},$$

or, with reference to Fig. 180,

$$\text{Elasticity} = \frac{P'M}{\dfrac{PM}{pP}}, \tag{1}$$

If PP' is produced until it meets the axis of pressure in E, we have two similar right angled triangles, namely, EpP and $P'MP$. Accordingly,

$$Ep : pP = P'M : PM,$$

or

$$Ep \, . \, PM = pP \, . \, P'M,$$

and

$$\therefore \quad Ep = \frac{pP \, . \, P'M}{PM}.$$

Dividing numerator and denominator of this fraction by pP, we have

$$Ep = \frac{P'M}{\dfrac{PM}{pP}}.$$

But from equation (1) above, we have seen that the expression forming the right hand member of this equation represents the elasticity. Therefore,

$$Ep = \text{elasticity}.$$

The numerator, $P'M$, of the above expression represents *a pressure*, while the denominator, $\dfrac{PM}{pP}$, denotes *the ratio of the change of volume to the original volume, and this ratio is simply an abstract number. The elasticity of a gas, therefore, is of the same nature as a pressure*, and the distance Ep measures the elasticity of the gas in terms of the same units in which the distance Op measures the pressure of the gas. Hence the practical rule for finding the elasticity of a gas at any point P of the isothermal curve is as follows: Draw the tangent PE to the isothermal curve at P, Fig. 181, and at the same time drop the perpendicular Pp from P upon the axis of pressures, then the intercept Op will represent the pressure of the gas under that condition, and pE will indicate the elasticity of the gas under the same condition.

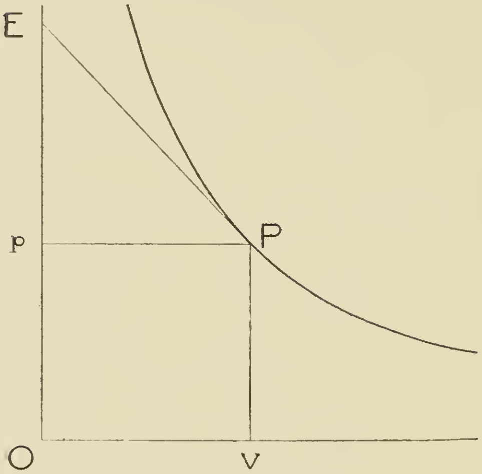

FIG. 181. The elasticity of a gas is numerically equal to the pressure.

There is another beautiful result which may be obtained from this diagram, owing to a remarkable property of the rectangular hyperbola. If from any point P on this curve, a tangent is drawn, and also from the same point a perpendicular is let fall upon one of the asymptotes, the foot of the perpendicular bisects the intercept produced by the tangent upon the asymptote. From this property of the rectangular hyperbola, it is seen in the diagram that the point p bisects the line $O\,E$, and therefore

$$pE = Op.$$

But pE represents the elasticity of the gas and Op the pressure. Therefore, in a gas which is compressed according to Boyle's law, whereby the condition of the gas may be represented by a point upon an isothermal curve, *the elasticity of the gas is numerically equal to the pressure.*

286. The Isothermals for Carbonic Acid Gas. If a mass of carbonic acid gas is enclosed in a cylinder and subjected to pressure by forcing in a piston, it will be found that, if the compression is carried on isothermally, the gas will obey Boyle's law almost exactly until the pressure has increased to twenty or thirty atmospheres. As the pressure, however, increases further, the slight deviations from this law, which have already been apparent, increase in magnitude, and the volume diminishes more rapidly than the pressure increases. If the isothermal compression is conducted at the temperature of 13° .1 C., the pressure will increase until it reaches 49 atmospheres, at which point the gas begins to liquefy. That is, the vapor of carbonic acid becomes saturated at this temperature and pressure. If the volume is now still further reduced, the pressure remains perfectly constant until the vapor has been entirely liquefied. After this, a further attempt to diminish the volume will cause the pressures to rise enormously, since the liquid is practicably incompressible. The form which the isothermal curve for 13°.1 assumes during the liquefaction of the gas is indicated in Fig. 182. Here, to save space, only the upper portion of the curve is shown, the true origin of pressures being about four inches below, so that the bottom of the figure represents a pressure of 45 atmospheres.

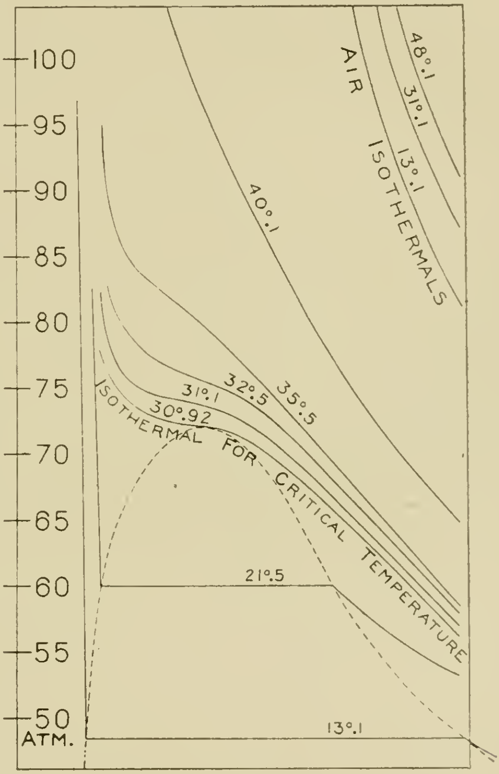

Fig. 182. Isothermals for carbonic acid.

If a similar compression is conducted at 21°.5, the gas will behave in an analogous manner, but the pressure will rise to 60 atmospheres before liquefaction occurs. This means that the saturated vapor of carbonic acid at 21°.5 exerts a vapor pressure of 60 atmospheres. Since the liquefaction now begins, the pressure remains perfectly constant at this figure until the vapor has all been reduced to the liquid state. As soon as this has been attained, the pressure rises very rapidly for a small decrease of volume, owing to the difficulty of compressing the liquid. The volume of the liquid carbonic acid at 21°.5 is greater than that at 13°.1 because of the large coefficient of expansion of this liquid. In this curve, as also in the isothermal for 13°.1, the portion traced during the liquefaction of the gas is horizontal, showing that the pressure remains perfectly constant during the diminuation of the volume of the saturated vapor.

If an isothermal curve is drawn for a temperature of 31°.1, it will be found that the form of this curve will be totally different from either of the others. The volume of the gas will diminish as the pressure increases until a pressure of 73 atmospheres is reached. The pressure then increases but slowly, while the volume diminishes very rapidly, but not the slightest trace of liquefaction is to be seen. A further reduction of the volume, however, causes an enormous rise of pressure, very much as if the effort were made to compress a liquid. The forms which the isothermals assume when the compression is carried on at 32°.5; 35°.5, and 48°.1 may be better understood from the figure than can be told in words.

287. The Critical Temperature. Dr. Andrews of England, who conducted the series of experiments just described, has shown that there is a certain temperature for carbonic acid above which no amount of pressure will condense the gas to the liquid form. This temperature, which is 30°.92, he called *the critical temperature;* and the corresponding pressure, 72½ atmospheres, he called *the critical pressure.* For all gases a similar critical point has been found, the temperature of which depends upon the nature of the gas. Accordingly, as has been defined in a former paragraph (p. 292), *the critical temperature of a gas is that temperature above which no pressure, however great, will reduce the gas to the liquid form.*

The isothermal of the critical temperature, Fig. 182, divides the plane of the diagram into two parts. The part to the right of this line may be called the gaseous region, because at a temperature above 30°.92 carbonic acid behaves exactly like a permanent gas, that is, above this temperature it cannot be liquefied by pressure. The region to the left of this isothermal may be divided into three parts. The part enclosed by the dotted line represents the gas as partly in the condition of a vapor and partly in the condition of a liquid. That is, within this region the saturated vapor can exist in contact with the liquid, the two being separated by a clearly-marked surface. The region to the right between the dotted line and the isothermal for the critical temperature may be called the region of the unsaturated vapor, because here the gas is below the critical temperature, and a decrease in the volume will produce an increase in the pressure. Finally, the portion of the diagram to the left of these two lines represents the region of the liquid.

This wonderful characteristic of the critical temperature has led Dr. Andrews to use this point as the feature which distinguishes a gas from a vapor. Inasmuch as the properties of many vapors can be determined only when they are in contact with their liquids, and since this is possible only at a temperature below the critical temperature, Dr. Andrews has defined a vapor to be a gas at a temperature below the critical temperature. To use his own words: "We may accordingly define a vapor to be a gas at any temperature under its critical point. According to this definition, a vapor may, by pressure alone, be changed into a liquid, and may therefore exist in the presence of its own liquid; while a gas cannot be liquefied by pressure—that is, so changed by pressure as to become a visible liquid distinguished by a surface of demarcation from the gas.".

Although the experiment is not unattended with the danger of an explosion, it is possible, by first guarding one's eyes with a sheet of glass, carefully to warm a sealed glass tube half filled with liquid carbonic acid to a point above the critical temperature for this gas. At ordinary room temperatures, the lower part of the tube is filled with the clear, colorless liquid, while the upper part is filled with saturated vapor, the two being separated by a distinct liquid surface with a well-defined meniscus. As the tube is warmed, the increased kinetic energy of the molecules causes the liquid rapidly and enormously to expand, and thus to decrease in density, while the more rapid evaporation and the decreased volume occupied by the vapor cause the latter greatly to increase in density. *At the critical temperature, the density of the vapor exactly equals that of the liquid,* so that the surface suddenly vanishes, and the distinction between the two disappears.

The tube, which is then above the critical temperature of the liquid, remains uniformly filled with the gas which no amount of pressure could ever reduce to the liquid form, even though the density were thus made many times as great as that of the liquid. This indicates that above the critical temperature the kinetic energy of the molecules is so great that, although these particles are forced by pressure very near together, it is impossible for the cohesive forces to act and form a liquid. As soon, however, as the temperature falls below the critical temperature, a white cloud is momentarily formed, within which the meniscus is seen to separate the liquid below from the saturated vapor above.

No energy change occurs at the critical temperature, such as otherwise attends the volatilization of a liquid or the condensation of a vapor.

288. Adiabatic Compression and Adiabatic Lines. In the compression of a gas, as we have heretofore considered it, we have assumed that the walls of the cylinder and also the piston are good conductors of heat. When, therefore, the temperature of the gas is raised by compression, heat will flow from the gas through the walls into the surrounding medium, which is kept at a constant temperature, and when the temperature of the gas is lowered by expansion, heat will flow from the surrounding medium into the gas; and hence we have assumed that however the pressure and volume may vary, *the gas always remains at a constant temperature.* This is what is meant by the *isothermal* compression or expansion of the gas.

It is possible, however, to compress or to expand a gas under conditions very different from those just considered, and these altered conditions give rise to an entirely new law governing the relation between pressure and volume. We will now assume the gas to be enclosed in a cylinder provided with a piston, as before, but in this instance we also assume that *the walls of the cylinder, as well as the piston, are absolute non-conductors of heat,* so that whether the temperature of the gas is high or low, *no heat can pass through the walls,* either from the gas to the medium outside, or from the outside medium to the gas within. If, therefore, the gas is compressed in such a cylinder, the heat produced by the compression remains in the gas itself, and so raises its temperature; and if the gas expands, the cold produced by the expansion lowers the temperature of the gas, and the latter remains at a low temperature because no heat can pass through the walls in either direction. The compression or expansion of a gas when enclosed in a cylinder of this nature, is known as *adiabatic.* The word adiabatic means "cannot pass through"; and therefore the adiabatic compression of a gas means a gas compressed in such a way that the heat produced by the compression "cannot pass through" the walls of the containing vessel.

When the volume of the gas is diminished, the pressure increases more rapidly for adiabatic compression than for isothermal. This may be seen if we assume for a moment that the compression takes place isothermally, in which case the point P, Fig. 183, will move to P' along the isothermal curve PP', and the volume Ov will be reduced to the volume Ov'. After this isothermal compression has taken place, we may think of the gas, which now occupies the volume Ov', as heated from without, until it acquires the same temperature which it would have had if the compression from the volume Ov to Ov' has taken place adiabatically. When now the gas is heated, the volume Ov' remaining constant during the heating, the pressure increases according to the law of Charles, and will

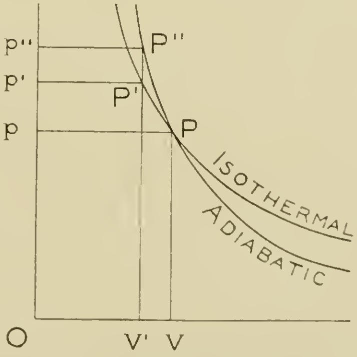

FIG. 183. Adiabatic and isothermal lines.

rise from the pressure P' to some other value such as P'', which is higher than P'. The gas then has the temperature which it would have had if the compression had taken place adiabatically: and hence, as the volume of the gas is reduced adiabatically from Ov to Ov', the pressure increases from Op to Op'', and the point P will trace out the curve PP'', which is a portion of the adiabatic curve. Since P is common to both curves, and since P'' on the adiabatic curve is higher than P' on the isothermal curve, it follows that the adiabatic curve is steeper than the isothermal.

We see, accordingly, that for a given diminution of volume, the pressure rises more rapidly for adiabatic than for isothermal compression, and therefore the elasticity of the gas (p. 310) is greater when the gas is compressed adiabatically than when it is compressed isothermally. *There are, accordingly, two elasticities of a gas, one for adiabatic compression and the other for isothermal compression.* Furthermore, it may be shown mathematically that the ratio of these two elasticities is the same as the ratio of their two specific heats (p. 262). That is,

$$\frac{\text{Elasticity under adiabatic compression}}{\text{Elasticity under isothermal compression}} = 1.41.$$

During adiabatic compression or expansion the gas does not obey Boyle's law.

289. The Work done in Compressing a Gas. We have seen that a single point in a diagram such as we have considered enables us to determine the condition of a gas as regards its volume and pressure, and that this point, as it moves, traces out a curve which is isothermal or adiabatic according to the nature of the compression or expansion. This same diagram enables us also to determine the work done in compressing a gas, as may be seen from the following consideration.

Let us suppose, as before, that the mass of gas is enclosed in a cylinder provided with a piston, and that the area of this piston is A, Fig. 184. Let us suppose also that the gas exerts a pressure of p units of force per unit of area within the cylinder, so that the total force upon the piston is Ap. If this piston is held in the cylinder by an external force P, then the force from without holds in equilibrium the force from within, and

$$P = Ap.$$

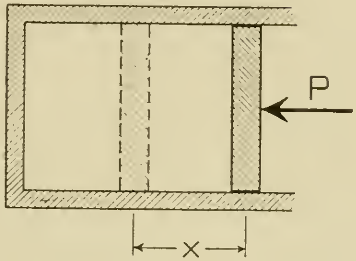

FIG. 184. Work is performed in compressing a gas.

As the piston is forced into the cylinder to compress the gas, P must increase and this causes p likewise to increase. In what follows, therefore, we shall assume P to be the average external force upon the piston and p to be the average pressure of the gas, that is, the average force per unit of area within the cylinder. If now the piston is pushed into the cylinder through the distance x, thus diminishing the volume, the work done by this compression is

$$\text{Work} = Px = Apx.$$

But A is the area of the piston and x is the distance through which the piston has moved, therefore Ax *represents the decrease in the volume of the gas*, because the volume of a cylinder equals the area of its base multiplied by its height. Letting V represent this decrease in volume during compression, we have

$$V = Ax.$$

Substituting this value for Ax in the above expression for work, we have

$$\text{Work} = Vp.$$

Therefore, *the work done in compressing a gas equals the product of the change of volume which has taken place and the average pressure of the gas during compression.* It is to be observed that this expression,

$$\text{Work} = Vp,$$

is entirely independent of the area of the piston and of the shape or size of the cylinder containing the gas, and depends solely upon the decrease in volume which has taken place and upon the average pressure exerted by the gas during compression.

The determination of the amount of work done during the compression of a gas becomes not only very evident, but very simple if we call to our aid the diagram of volumes and pressures heretofore considered. Again, if the horizontal line represents the axis of volumes and the vertical line the axis of pressures, Fig. 185, the point P indicates that the gas at that instant has a volume represented by Ov and a pressure represented by Op, or its equal vP. If the gas is now compressed so that it occupies the volume represented by Ov', the change in volume is indicated by

$$Ov - Ov' = vv'.$$

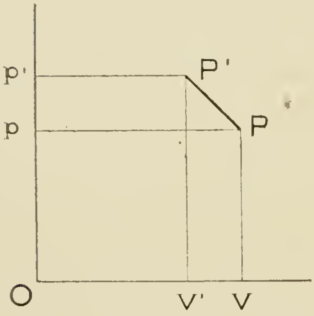

FIG 185. Diagram showing the the work done in compressing a gas.

During this change, the point P has moved from P to P' along a curve which may be either isothermal or adiabatic, where PP' is a portion of this curve so small that it may be considered straight. The pressure, in the meanwhile, has increased from Op to Op', the average pressure being

$$\frac{Op + Op'}{2},$$

or, what is the same thing,

$$\frac{vP + v'P'}{2}.$$

We have seen above, however, that the work done in compressing a gas is equal to the change of volume multiplied by the average pressure, so that

$$\frac{vP + v'P'}{2} \cdot vv'$$

represents the work which has been done during this compression. But

$$\frac{vP + v'P'}{2}$$

is the average height of the figure *vv′P′P,* and

$$\frac{vP + v'P'}{2} \cdot vv'$$

is therefore also the area of this quadrilateral. The work done, therefore, in compressing the gas from the volume *Ov* to the volume *Ov′*, while the pressure changes from *vP* to *v′P′*, is represented by the area *vv′P′P*. If on the other hand, we consider that the gas expands from the volume *Ov′* to the volume *Ov*, while in so doing the pressure falls from *v′P′* to *vP,* then the gas does an amount of work represented by the area *vv′P′P*.

It is to be observed that if the point *P* moves to the right as it traces its curve in the diagram, the gas itself does the work, and the work thus done by the gas may be called positive. If, however, the point *P* moves to the left, then it is the outside force that compresses the gas and does the work, and the work thus done upon the gas may be called negative. For example, referring to the last diagram, if the gas expands from the volume *Ov′* to *Ov*, thus pushing back the external pressure, the pressure of the gas changing from *v′P′* to *vP* in so doing, the gas itself performs an amount of work represented by the area *vv′P′P*, the point *P* moves to the right, and this amount of work is positive. If, on the other hand, the outside force compresses the gas, changing its volume from *Ov* to *Ov′* while the pressure rises from *Op* to *Op′*, then the outside force does an amount of work upon the gas indicated by the area *vv′P′P*, the point *P* moves to the left, and this amount of work is negative. An application of this important principle will be made in the consideration of Carnot's cycle (p. 325).

290. The Mechanical Equivalent of Heat—The First Law of Thermodynamics. The investigations of Count Rumford and Sir Humphrey Davy (p. 231) established beyond a doubt the fact that heat is a form of energy. It remained, however, for James Prescott Joule to make an absolute determination of the relation between the quantity of heat developed and the mechanical energy expended. Joule's apparatus consisted of a falling weight, Fig. 186, which set in motion a paddle-wheel in a cylinder filled with water. In this experiment the energy represented by the falling weight was consumed in churning the water within the cylinder, and this produced a rise of temperature of the water. After many years of the most zealous labor, Joule announced as the result of these investigations that the work done by the mass of one pound falling through the distance of 772 feet, when transformed into heat, will raise the temperature of 1-pound of water 1° F. Expressed in terms of the modern units of measurements, this is the equivalent of expending 4180 joules of work to develop 1 larger calorie of heat, the joule being thus named in honor of this great investigator who conceived and carried out this remarkable research.

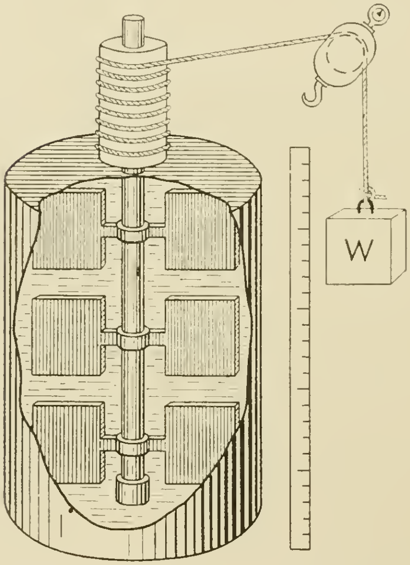

Fig 186. Joule's device for determining the mechanical equivalent of heat.

While these results were of the greatest importance, it was recognized that, inasmuch as Joule worked with a comparatively small amount of water, and that this was raised only a single degree F., the exact figure which he announced, expressing the relation between energy expended and heat developed, might be somewhat in error. The fact that later experimenters with improved apparatus obtained values which are almost identical with this points to the infinite care and devotion with which Joule conducted his work.

Later, Rowland of Johns Hopkins University, made a redetermination of the mechnical equivalent of heat, working with a larger quantities than Joule and taking greater precautions. Rowland drove his paddle-wheel by a steam engine and determined the amount of work done by measuring the moment of the couple required to prevent the cylinder containing the water from turning. At the conclusion of these experiments, which were conducted with the utmost accuracy, he announced, as the result, that to develop one calorie of heat requires the expenditure of 427 kilogrammeters of work. This result, namely, that one calorie of heat is the equivalent of 427 kilogrammeters, or 4187 joules, of mechanical energy expended is called the mechanical equivalent of heat, or Joule's equivalent.

The fact that a perfectly definite relation exists between the quantity of heat developed and the amount of mechanical energy expended lies at the basis of all our modern conceptions regarding the nature of heat. Indeed, this fact has been so firmly established that the statement of this truth is now called the first law of thermodynamics; and may be expressed in the following words: *To a quantity of heat developed there corresponds a perfectly definite amount of mechanical energy expended.*

291. Carnot's Cycle. We shall now examine the action of an ideal engine in which the material upon which the various operations are performed may be considered as a gas, and the expansions and compressions of this gas are supposed to take place isothermally or adiabatically. This engine is wholly ideal in its action; but, although it cannot be constructed, it is, nevertheless most useful in giving information regarding the dynamical theory of heat. All parts, therefore, may be imagined to be created for the pure purpose of explanation. Its conception is due to Carnot, and the various processes herein employed are known as Carnot's cycle of operations, or Carnot's engine.

In the study of Carnot's cycle, we assume a mass of gas enclosed in a cylinder provided with a piston, Fig. 187, but the walls of the cylinder and also the piston are assumed to be absolutely imprevious to heat, that is, perfect non-conductors of heat, while the bottom of the cylinder is assumed to be a perfect conductor of heat, so that the enclosed gas will always be at the temperature of the surface upon which the cylinder is placed. We further assume that we have two indefinitely large reservoirs of heat, one, called the hot body, being maintained at the absolute temperature T, and the other, called the cold body, or the refrigerator, being maintained at the absolute temperature t. In other words, we assume these bodies to be so large that if a small quantity of heat is added to them or taken from them, their temperatures will not be materially changed. More-

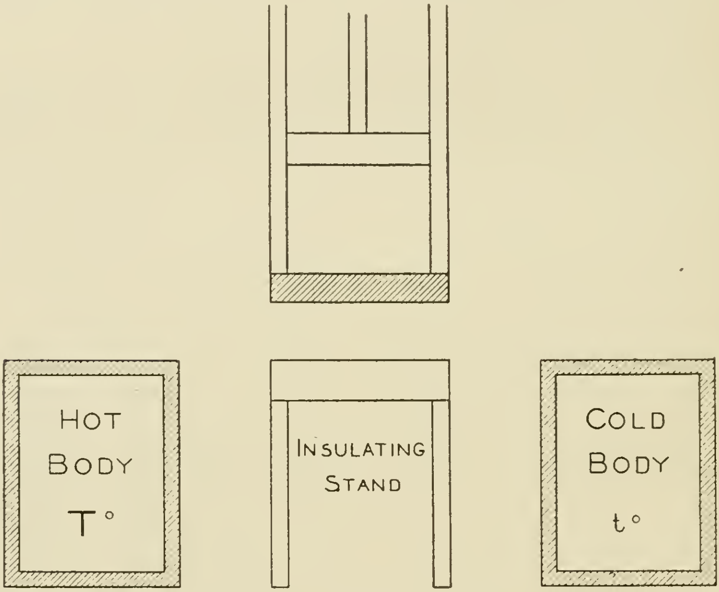

FIG. 187. Carnot's engine.

over, we assume the existence of a third body which is a perfect insulator for heat so that no heat may be conducted either to or from the cylinder of gas when placed upon it. We may, indeed, assume that this third body is an insulating stand, of such nature that when the cylinder is placed upon it, the gas is completely cut off from a flow of heat.

The meaning of the word cycle as applied to the series of operations which we are about to describe means that the gas in the cylinder is to be carried through a number of operations, and is finally to be brought back again, so that the condition of the gas will be in every way the same at the end of these operations as it was in the beginning. This cycle consists of four separate operations which must be carried out *in order,* any one of which, however, may be taken as the first one. We shall, therefore, begin this cycle as follows:

Operation 1. The cylinder containing the gas at the temperature t of the cold body is placed upon the insulating stand and the piston pressed down, thereby compressing the gas. Inasmuch as no heat can flow through the conducting bottom of the cylinder, since this is on the insulating stand, the heat produced by the compression remains in the gas, and the point P, Fig. 188, which indicates the condition of the gas as the cylinder was placed on the insulating stand, now moves along the adiabatic line PQ. The adiabatic compression is now continued until the temperature of the gas rises to the temperature T of the hot body. In producing this compression, the outside force which pressed the piston down has performed upon the gas an amount of work equal to $PQba$, and since the point in the diagram moves to the left, this may be considered negative work

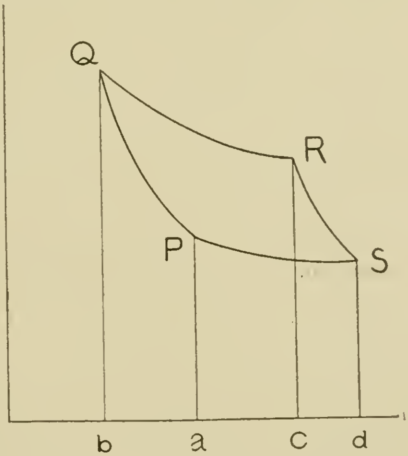

FIG. 188. Carnot's cycle.

Operation 2. The cylinder is then removed from the insulating stand and placed upon the hot body and the piston allowed to rise. The gas then expands, and an amount of heat equal to H at the temperature T flows into the gas through the conducting bottom of the cylinder. As the gas expands, the moving point in the diagram moves to the right along the isothermal curve QR and the gas, therefore, does an amount of positive work equal to the area $QRcb$.

Operation 3. The cylinder is now placed upon the insulating stand, and the gas allowed to expand adiabatically until the temperature falls to that of the cold body, t. During this process the moving point in the diagram describes the adiabatic curve RS; and since the point moves to the right, the gas does an amount of positive work represented by the area $RSdc$. Moreover, this work is done at the expense of the heat which the gas contained, hence the temperature falls.

Operation 4. The cylinder of gas, which is now at the temperature t, is placed upon the cold body, and the piston forced down. The gas is then compressed isothermally at the temperature of the cold body until it has again attained its original volume indicated by P in the diagram, and an amount of heat equal to h at the temperature t is forced from the gas into the cold body. The moving point in the diagram moves to the left along the isothermal SP, and indicates that an amount of negative work equal to the area $SPad$ has been done by the outside force upon the gas.

At the end of these four operations the gas is in exactly the same condition as at the beginning, for it once more has the temperature t of the cold body, and its volume and pressure are again indicated as before by the point P in the diagram. During the second operation, however, a quantity of heat equal to H, at the temperature T, flowed from the hot body into the gas, and in the fourth operation a quantity of heat equal to h, at the temperature t, passed from the gas into the cold body. During the first and third operations, the gas neither received nor gave up heat. Moreover, during the second and third operations the gas performed an amount of positive work equal to the area $QRSdb$, while during the first and fourth operation the outside pressure did an amount of negative work upon the gas equal to the area *SPQbd*. Accordingly, the amount of work performed by the gas in excess of the work performed upon it is the area of the figure *PQRS*.

The results accomplished by this series of four operations upon the gas are three in number, as follows:

1. An amount of heat equal to H at the temperature T has been drawn from the hot body.

2. An amount of heat equal to h at the temperature t has been delivered to the cold body.

3. An amount of mechanical work equal to the area of the figure *PQRS* has been performed.

It is to be observed that two results are produced by the quantity of heat H which is drawn from the hot body:

1. A quantity of heat equal to h has been transferred to the cold body.

2. A quantity of mechanical work equal to *PQRS* has been performed.

Therefore, a quantity of heat has passed from a higher temperature to a lower temperature, and an amount of mechanical work has been performed.

292. The Indicator Diagram and the Efficiency of the Steam Engine. In practice this theoretical cycle of Carnot finds its application in the indicator card, or indicator diagram, which may be taken from any actual steam engine in operation, and which shows the amount of work the engine is doing. In taking this diagram, a pencil point is made to move automatically in the horizontal direction by the stroke of the piston, and, at the same time, it is actuated in the vertical direction by the pressure of the steam in the cylinder. Distances in the horizontal direction, therefore, are proportional to the volume of the steam in the cylinder, while distances in the vertical direction are proportional to the varying pressure of the steam. After one complete stroke of the engine forward and back, the steam in the cylinder is again in its initial condition, so that the pencil point has traced a closed curve, Fig. 189, which roughly resembles the figure *PQRS* of Carnot's cycle. While the steam ports are open, the expansion or compression of the steam takes place nearly isothermally, and when the steam ports are closed, these changes occur approximately adiabatically. The area of the diagram is then carefully measured, and from this is computed the work which the engine is doing.

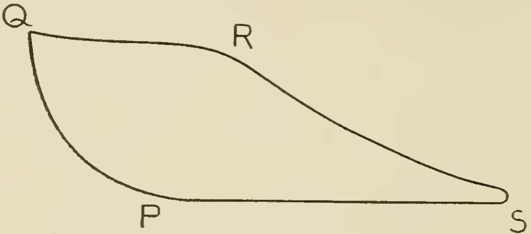

Fig. 189. A steam engine indicator diagram.

In the ordinary high pressure steam engine, the hot body is the boiler, and its temperature, T, is the temperature of the steam as it is delivered to the cylinder; while the cold body is the air, into which the exhaust steam is forced, and its temperature, t, is 100° C., which is the temperature at which steam can exist at atmospheric pressure. Since the quantity of heat H is taken from the hot body, or the boiler, and since the quantity h is given to the cold body, that is, forced out into the air, it follows that the difference between these two, namely,

$$H - h$$

has been converted into work. Moreover, the ratio of this amount of heat which has been converted into work to the quantity originally taken from the hot body, that is,

$$\frac{H - h}{H}$$

represents the efficiency of the engine (p. 123). Furthermore, it may be shown mathematically that the quantity of heat H is proportional to the absolute temperature T of the hot body, while the quantity of heat h is proportional to the absolute temperature t of the cold body, so that

$$\frac{H-h}{H}=\frac{T-t}{T}.$$

This, however, is true only in the theoretical reversible engine. In the practical engine

$$\frac{H-h}{H}<\frac{T-t}{T}.$$

It is now easy to see how very low the efficiency of even the best steam engine really is. Under a pressure in the boiler which it is safe to carry, the steam rarely exceeds a temperature of 165° C., or 438° absolute, while the temperature at which it exhausts into the open air is 100° C., or 373° absolute. When these values are substituted for T and t, respectively, it is then seen that the theoretical efficiency of a steam engine under normal conditions can hardly exceed 15 per cent. By taking the precaution to superheat the steam to 206.3° C., and by condensing the steam in a chamber at 43.1° C., instead of allowing it to exhaust into the air at 100° C., the manufacturers of the Nordberg quadruple expansion engine are able to obtain an actual efficiency of 25.5 per cent, and a theoretical efficiency of 34 per cent.

Moreover, a low efficiency is inherent in the nature of the steam engine, for the efficiency of any engine depends solely upon the absolute temperatures of the two bodies between which it works. The nearer the temperature of the cold body approaches the temperature of the absolute zero, the higher becomes the efficiency of the engine.

293. The Reversibility of Carnot's Cycle, and the Second Law of Thermodynamics. One further beauty of Carnot's cycle is the fact that it may be worked backward, and the several operations may all be made to take place in the reverse order. To show this, let us again suppose the gas to be in the cylinder in the condition represented by the point P, Fig. 190, and at a temperature t of the cold body.

Operation 1. The cylinder is placed on the cold body and the gas allowed to expand along the isothermal *PS*. During this expansion an amount of heat equal to h passes through the conducting bottom of the cylinder into the gas within, and the gas performs an amount of positive work equal to *PSda*.

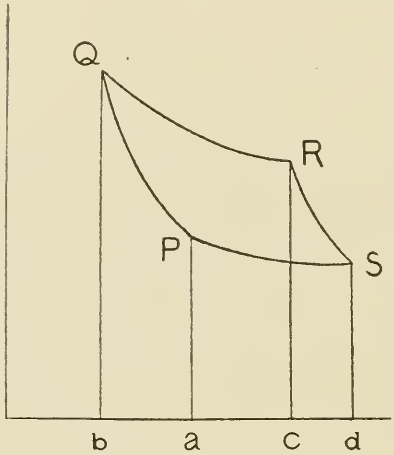

Fig. 190. Carnot's cycle—the engine operating backwards.

Operation 2. The cylinder is then placed on the insulating stand and the gas compressed along the adiabatic line *SR* until the temperature has risen to the temperature T of the hot body. During this adiabatic compression the outside force performs an amount of negative work equivalent to *SRcd*.

Operation 3. The cylinder is now placed on the hot body and the gas compressed along the isothermal curve *RQ*. During this compression, a quantity of heat equal to H is transferred to the hot body and the outside pressure performs an amount of negative work equivalent to *RQbc*.

Operation 4. The cylinder is finally placed upon the insulating stand and the gas allowed to expand adiabatically along the line *QP*, the gas therefore performing an amount of positive work equal to *QPab*.

During these four reverse operations, the gas in expanding has performed an amount of work equal to *QPSdb*, while the outside force in compressing the gas has performed an amount of work equal to *SRQbd*. The work, therefore, which the outside force has done in excess of the work which the gas has done is the area of the figure *PQRS*.

Accordingly, the results of these four reverse operations are three in number:

1. A quantity of heat h at the temperature t has been taken from the cold body.

2. A quantity of heat H at the temperature T has been transferred to the hot body.

3. The outside force which compressed the gas has performed an amount of work equal to $PQRS$.

The great points to be observed in these two cycles of operations, direct and reverse, are the following:

1. That *heat will pass of itself from a hot body to a cold body and perform work.*

2. That *heat will pass from a cold body to a hot body only if external work is done upon it.*

The truths embodied in these two statements form the great second law of thermodynamics, and this all-important law may, indeed, be expressed in the words here given. It is more usual, however, to state this second law in the following form: *Heat cannot pass without compensation from a body of lower temperature to a body of higher temperature.*

STATIC ELECTRICITY

CHAPTER XXII.

THEORIES AND PHENOMENA.

294. Early Discoveries. About six hundred years before Christ, the old Greek philosopher, Thales, discovered that amber when rubbed with silk acquires the power of attracting light bodies, such as bits of chaff and straw and particles of dust. For more than two thousand years this single fact was all that was known regarding this phenomenon, and it was supposed that amber was the only substance which could acquire this remarkable property. Indeed, it was not until the year 1600 that Gilbert, who was physician to Queen Elizabeth, discovered that other substances, such as sulphur, glass, and all resinous bodies, upon being rubbed, acquire this wonderful new power possessed by amber. Inasmuch as the Greek name for amber is *elektron,* the name "electrics" was given to the whole class of bodies which in common with amber possess this property. The word electricity was adopted soon afterwards to indicate the "fluid", as it was then supposed to be, by means of which this effect was produced in many otherwise inert substances, and the term electrification is still used to denote the condition of a body after this new property has been imparted to it.

295. Later Investigations. Two generations after Gilbert's investigations, Robert Boyle, in 1662, discovered that there is a *mutual* attraction between the electrified body and the body attracted. This he proved by suspending the electrified body in such a way that it could turn freely, and observing that any body presented to it would attract it, just as it in turn would attract a light body. This is simply an illustration of the great truth which Newton saw and announced later (p. 107), that a single force is only one of a pair of two equal and opposite forces, or that action and reaction are equal and oppositely directed.

Boyle experimented with many other substances besides those tried by Gilbert, and found that the metals alone seemed to form an exception to the general rule that bodies become electrified by friction, for these apparently could not be electrified. Stephen Gray, in 1736, however, discovered that the metals form no exception, for when they are held by a handle formed of an "electric" while being rubbed, they also become electrified. The conclusion was then apparent that when a body is held in the hand while being rubbed, this condition, which remains upon "electrics", escapes from the metals, being conducted away through the hand, but that if a metal is provided with a handle of glass, sulphur, or resin, this escape is prevented.

Thus early was it seen that all bodies may be roughly divided into two great classes, namely, those which lose all traces of electrification upon being touched, and those which retain their electrification when held in the hand. It is thought that the former allow the electricity to pass or to "flow" through their substance, while the latter in some way prevent this passage. Bodies of the former class, like the metals, are therefore called conductors, while bodies of the latter class, such as glass, shellac, hard rubber, sulphur, amber, sealing-wax, flannel, silk, etc., are called non-conductors, or insulators. This distinction, however, cannot be too closely drawn, because there is no conductor, however good, that does not offer some opposition to the passage of electricity, and on the other hand there is no insulator so perfect that does not allow some slight passage of electricity through its material. The terms conductor and insulator, therefore, are wholly relative.

Towards the middle of the eighteenth century, interest in the new study of electricity was active, and progress was comparatively rapid. In 1733 Du Fay investigated the effect which one electrified body has upon another, and discovered that this is not always the same. We may repeat his general line of experiments by bringing one electrified hard rubber rod near another electrified hard rubber rod so suspended as to turn freely, Fig. 191, when it will be found that the two will repel each other. The same will be true if the experiment is tried with two electrified shellac rods, two electrified sulphur rods, two electrified glass rods, etc. In fact, any two bodies of the same kind that have been electrified in the same way will repel each other.

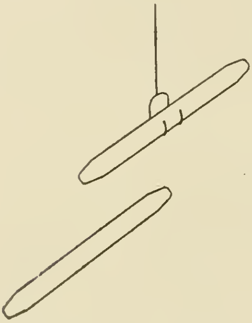

Fig. 191. Two similarly electrified bodies repel each other.

Moreover, it will be found that a suspended electrified rod of hard rubber will also be repelled by an electrified rod of sulphur, sealing-wax, shellac, amber, or any resinous substance, but that it will be attracted by an electrified rod of glass. This shows that the electrification of sulphur, sealing-wax, shellac, amber, and all resins is the same as that of hard rubber, and that this electrification is of a different nature from that produced upon glass.

The all-important fact which Du Fay made evident is that *there are two different kinds of electrification*, one produced upon glass when rubbed, and the other developed upon bodies of a resinous nature. The kind produced upon glass he called the vitreous electrification, while that which appears upon the resins he called the resinous.

296. Mutual Action of Electrified Bodies. The mutual attractions and repulsions between electrified bodies give rise to the following two great laws, which are of fundamental importance, and which will be used in all our future study of this subject:

1. Two bodies similarly electrified repel each other.
2. Two bodies dissimilarly electrified attract each other.

It should be pointed out in this connection that the nature of the rubbing material influences the character of the electrification of the rubbed body, that is, a body which is vitreously electrified when rubbed with one substance often becomes resinously electrified when rubbed with another. A glass rod, for instance, becomes vitreously electrified when rubbed with a piece of silk but resinously electrified when rubbed with a cat's fur. Moreover, the nature of the surface of the rubbed body affects the kind of electrification produced. Thus, ordinary glass when rubbed with silk becomes vitreously electrified but ground glass rubbed with silk becomes resinously electrified. We can, however, depend upon the fact that a rod of smooth glass when rubbed with silk becomes vitreously electrified, and that a rod of hard rubber, shellac, sealing-wax, sulphur, amber, etc., when rubbed with flannel or a cat's skin becomes resinously electrified

297. Electroscopes. An electroscope is any device that enables us to determine whether or not a body is electrified. In this sense of the word, the suspended stirrup containing an electrified rod (p. 334) is an electroscope, because, if another similarly electrified body is approached, the two will repel each other. The suspended electrified rod, however, possesses large mass and large inertia, so that if a very small electrified body were approached, or if one very feebly electrified were brought up, the great inertia of the suspended rod would render its motion difficult of detection. While the above apparatus, therefore, forms an electroscope, it is not only inconvenient, but is far from delicate, owing to the great inertia of its moving part.

298. The Pith-Ball Electroscope. A much more convenient and sensitive device for detecting the presence of electricity upon a body is the pith-ball electroscope. This little instrument consists of two balls of pith, covered with gold-leaf to make them conductors, and suspended by delicate silk threads, Fig. 192. When these little balls of pith touch an electrified body, some of the electricity upon the body passes or flows off upon the little balls, thus electrifying, or *charging*, them with the same electrification as that upon the electrified body. These two pith-balls now being similarly electrified repel each other, and so do not lie side by side as when unelectrified, but are held a little distance apart by the remarkable force which exists between electrified bodies.

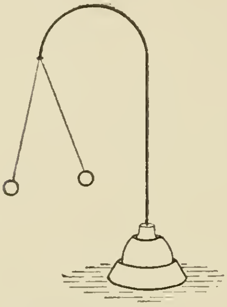

FIG. 192. The pitch ball electroscope.

Although a simple inspection of the instrument will thus enable the observer to decide whether or not the pith-balls are charged, the question as to the nature of their electrification cannot be so readily answered. To decide this matter a charged body must be approached, the electrification of which is known, and the effect upon the pith-balls observed. If the pith-balls are repelled by an electrified glass rod brought near them, the conclusion is obvious that the electrification of the pith-balls is vitreous, because bodies similarly electrified repel each other. If attraction instead of repulsion takes place, the inference must not be drawn that the electrification of the balls is resinous, because it has been seen (p. 332) that even an unelectrified, or neutral, body will attract an electrified body. A resinous body, such as hard rubber or sealing-wax which has been rubbed with flannel or cat's skin must then be approached, and if the little pith-balls are then repelled, it is certain that they are charged resinously.

This is but a particular case of a very general rule which must always be observed in testing the nature of the electrification. The nature of the electrification of the electroscope must be determined in the above mentioned way and the body, the electrification of which is to be determined, is then approached. If a repulsion of the moving parts occurs, it follows that the electroscope and the body are similarly electrified. If, however, attraction occurs, the electroscope must be oppositely charged and the test repeated. When repulsion occurs between the moving parts of the electroscope and the body under test, the two are then similarly electrified.

299. The Gold-Leaf Electroscope. There is, however, another form of electroscope, known as the gold-leaf electroscope, the moving parts of which possess less inertia and less weight than the pith balls, and which accordingly is more sensitive. This consists of a case with sides of glass enclosing two delicate strips of gold-leaf which are hung from the lower end of a metal rod, Fig. 193. This rod passes through the case, from which it is carefully insulated, and terminates above in a knob, or ball. When unelectrified, the gold-leaves hang together; but if the knob of the electroscope is touched by an electrified body, electricity flows from the body to the ball, then passes down the rod and distributes itself over the two gold-leaves. While these, accordingly, become similarly charged and repel each other, the nature of their electrification must then be tested, as was done with the pith balls in the pith-ball electroscope. An electrified glass rod must be approached, and if the leaves diverge still more widely, they are charged vitreously. If they tend to contract, a resinously electrified rod must be approached, when a further divergence of the leaves will follow. With the gold-leaf electroscope, as with the pith balls, the only sure indication of the nature of the electrification is a further divergence of the moving parts, for then it is certain that the electrification of the electroscope is of the same nature as that of the body tested.

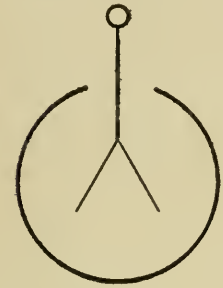

FIG. 193. The gold leaf electroscope.

300. The Law of Attraction and Repulsion. At times the leaves of the electroscope are observed to be violently affected, and at other times they can scarcely be seen to move, depending upon the degree of the electrification of the body which is approached. We assume this difference in the intensity of electrification to be due to a larger or smaller *quantity of electricity* residing upon the electrified body, a large quantity producing a greater effect than a smaller amount. Accordingly, inasmuch as we have to deal with quantities of electricity, we must select a unit, and the measurement of any given quantity must then follow in terms of the unit so chosen.

By the use of the torsion balance (p. 144), which he rediscovered, Coulomb found that the force of attraction or repulsion between two electrified bodies varies directly as the product of the quantities of electricity, and inversely as the square of the distance between them. That is, if F is the force of attraction or repulsion between two small spherical bodies, charged with quantities of electricity which may be indicated respectively by Q and Q', and if r is the distance between the centers of these bodies, then, if the electrified bodies are in the air,

$$F : \frac{Q\,Q'}{r^2},$$

or

$$F = k\,\frac{Q\,Q'}{r^2},$$

where k is the factor of proportionality.

From this expression we derive the definition of the unit quantity of electricity which is in common use.

301. The Unit Quantity of Electricity. Since we are free to choose any desired quantity of electricity as the unit, we may select for this purpose a quantity of such magnitude that both Q and Q' shall be unity when F and r are each unity, in which case,

$$k = 1,$$

and the above equation becames

$$F = \frac{Q\,Q'}{r^2}.$$

In thus giving to k the value unity, we have, however, determined the absolute unit quantity of electricity to be that quantity which at unit distance will repel a similar and equal quantity with unit force. Or, in terms of the units of the C. G. S. system (p. 18), *the absolute C. G. S. unit quantity of electricity is that quantity which, at the distance of 1 centimeter, will repel a similar and equal quantity with a force of 1 dyne.* This is known as the C. G. S. electrostatic unit of quantity.

302. The Dual Nature of Electrification. The great fact must now be made prominent that whenever one body is electrified in any way whatever, there exists somewhere another body which is also electrified. In other words, there are always, without exception, two bodies which undergo electrification at the same time. Moreover, if one of these bodies is electrified vitreously, the other is electrified resinously, and the quantity of electricity with which one is electrified just equals the quantity with which the other is electrified. For instance, when a glass rod is electrified vitreously with a certain quantity, the silk with which it was rubbed is found to be electrified resinously with the same quantity. So, also, when a rubber rod is electrified, resinously, the cat's fur with which it was rubbed is found to be electrified vitreously. Furthermore, the quantity with which the rubber rod is electrified just equals the quantity with which the cat's fur is electrified.

303. Early Theories of Electricity. Facts such as these, all of which point to the dual manifestation of electricity, have given rise to much speculation as to its nature. Du Fay conceived the idea of two electric fluids existing in all bodies, the relative amounts of which might to some extent be varied, and which completely neutralized each other when present in equal amounts in the same body. He thought that an excess of one fluid or the other produced electrification; and, as there are two forms under which the electrified condition appears, he assumed that the vitreous electrification was produced by an excess of one of these fluids, which he called the vitreous, and that the resinous electrification was due to an excess of the other fluid, which he termed the resinous. Du Fay's conception has received the name of the two-fluid theory.

In 1747 Benjamin Franklin propounded what is known as the one-fluid theory. This was based upon the assumption that there was a universal medium, or fluid, called electricity, which was present in all bodies, and which, under ordinary circumstances could not be detected, because all other bodies contained exactly equivalent amounts. If, however, the quantity present in a body were in any way changed, that is, if this natural or normal amount were either increased or decreased, the body became at once electrified. He assumed, furthermore that the vitreous electrification was due to an excess of this electricity above the normal amount, and to this he therefore gave the name positive; while the resinous was caused by a deficit in the normal amount, that is, that the electrified body had lost some of the electricity which it originally had, and this he called negative. Accordingly, the terms positive and negative were used exactly as they are used in algebra, not to indicate a difference in the *kind* of electricity with which a body was charged, but simply to show whether that body contained a quantity greater or less than the normal amount.

Each of these theories explained the observed phenomena perhaps equally well. Within the last few years, however, since the discovery and isolation of the electron, both of these older views have given way to the newer aspect which the entire science of electricity has assumed.

304. The Nature of the Electrons. As has already been pointed out (p. 11), the atoms of matter are believed to consist, in part, at least, of numbers of minute particles, called the electrons, revolving within the atoms with speeds which are but little less than the velocity of light.

Occasionally it transpires that an atom becomes unstable in such a way that an electron escapes from the miniature solar system in which it has been revolving, and thus is able to move about freely among the molecules of matter. This particle is then found, not only to retain the high velocity with which it left its parent atom, but always to be negatively electrified, and to be charged with a quantity of electricity equal to 4.77×10^{-10} electrostatic units. Since these electrons may be obtained from matter of all kinds, and since their properties are independent of the nature of the atoms from which they are derived, it is thought that these small particles form an integral part of all atoms, and therefore of all matter.

The further point, however, must be considered, that, if the atom were to consist only of these electrons, it would be unstable, since these negatively electrified particles would mutually repel each other. We believe, therefore that the atom must contain, besides the electrons, also a quantity of positive electricity equal in amount to the combined quantity carried by all the electrons. In order that the latter particles may remain in stable equilibrium within the atom, we conceive that they revolve, either about a positive nucleus at the center, or in a medium of ''positive electrification'' distributed uniformly throughout the atom. Indeed, it is invariably found that the escape of an electron from an atom leaves the latter positively electrified.

Moreover, when electrons move between two oppositely charged plates, they are repelled by one and attracted by the other, so that they are accelerated, and therefore move with a higher velocity. Their mass is then found to be not constant but variable, and to become greater when their velocity increases. Since their mass is not constant, but depends upon the velocity with which they move, the electrons can therefore not be matter in the ordinary sense, inasmuch as matter presupposes the constancy of mass. We believe also, therefore, that the electrons are purely electrical in their constitution; and since they always represent exactly the same quantity of electricity, wherever they are found and from whatever substance they may be derived, we believe they are the ultimate particles or units of electricity, and that electricity, in all of its manifestations, is due simply to the presence, or absence, or motion of these electrons.

We must therefore think of electrons as existing under two entirely different conditions: first, as revolving, probably in regular systems, within the atoms; and second, as freely-moving, independent particles distributed among the molecules of matter, bombarding and colliding with the latter, in their rapid motion, and being in turn bombarded by them. While thus moving freely among the molecules of matter, with which they are supposed to be in thermal equilibrium, they are known as the *free* electrons.

305. The Modern Theory of Electrification. At first it might seem that all bodies which contain large numbers of free electrons must be negatively electrified. It must be remembered, however, that the atoms from which these electrons were derived were originally neutral, and that the removal of an electron from an atom leaves the remainder of the atom charged positively. There are, therefore, in the body as many positively charged particles as negative, and the body, as a whole, is still neutral.

By a number of processes, it is found possible to alter the number of electrons in a body or upon its surface. Prominent among these methods is friction, or the rubbing of one substance upon another. From this point of view, the electrification which results when a glass rod is rubbed with silk is produced by a redistribution of the electrons resident upon the two materials, whereby a number of electrons are transferred from the surface of the glass to the silk. This removal of a number of these negatively charged electrons from the glass leaves the latter charged positively; while, on the other hand, the addition to the silk of the electrons which were thus taken from the glass gives to the silk an excess of these particles, which thereby charges it negatively. The silk is therefore not only charged negatively, but is charged negatively to a degree precisely as great as that to which the glass is charged positively. In a similar manner, a rod of hard rubber or any one of the resins, when rubbed with flannel or cat's fur, becomes charged negatively, inasmuch as in this instance a number of the electrons are transferred from the flannel or the fur to the rod, thus charging the latter negatively, and leaving the flannel or the fur charged positively to an equal degree. Even the separation of any two dissimilar substances will bring about a redistribution of the electrons common to the two.

Accordingly, when two bodies are electrified by friction, for two must always be electrified equally and oppositely at the same time, one will carry with it a number of electrons greater than it formerly had, and will therefore be negatively electrified, while the other will be left with less than its normal number of these particles, and will be positively electrified. A negatively charged body, then, is one upon which there is an excess of electrons, while a positively charged body is one which has lost some of the electrons which were normal to it and which it originally had.

It is thus seen that, except for the sign of the charge, the modern theory of electrification is strikingly similar to the one-fluid theory, as proposed by Franklin. Indeed, if Franklin had chanced to give exactly the opposite signs to his positive and negative electrifications, it would have simplified many matters connected with the flow of electricity as we now understand it. What we mean by a flow of electricity through a body is simply a directed movement of the free electrons in that body. The distinction between a conductor and an insulator is thus seen to depend entirely upon the relative number of free electrons that may be present, a conductor being a body which contains large numbers of free electrons, while an insulator is one having very few of these particles or practically none.

306. Surface Density. When a conductor is charged with electricity, it is thought that a certain number of units of electricity are distributed over its surface, and the surface density of the charge, which is usually denoted by the Greek letter σ, is the number of units of electricity upon one square centimeter of the surface. Letting S represent the number of square centimeters on the surface of a body, and Q the number of units of electricity uniformly distributed over the surface of that body, we see that

$$\frac{Q}{S}$$

represents the number of units of electricity upon one square centimeter, or σ, and therefore

$$\sigma = \frac{Q}{S}.$$

If the distribution of electricity is uniform, as it is over the surface of a charged insulated metal sphere, then σ is constant for every part of the surface; but if the distribution is not uniform, then σ varies from place to place on the body.

307. Surface Density on a Sphere. We have seen that the surface density of the charge is the number of units of electricity upon one square centimeter of the surface. If, therefore, the charge of Q units of electricity is uniformly distributed over the surface S of an insulated metallic sphere, the surface density at every point is

$$\sigma = \frac{Q}{S}.$$

If R is the radius of the sphere, the surface S represents an area of $4 \pi R^2$ square centimeters, that is,

$$S = 4 \pi R^2,$$

and accordingly the surface density for a sphere becomes

$$\sigma = \frac{Q}{4 \pi R^2},$$

or the quantity of electricity upon a charged insulated sphere is

$$Q = 4 \pi R^2 \sigma.$$

308. The Distribution of Electricity upon a Conductor. The distribution of electricity upon any surface may be determined experimentally by the device known as the proof-plane. This consists of a small metal disc mounted upon an insulating handle, or a small metal sphere suspended by a silk thread, Fig. 194. The proof-plane may be touched to an electrified body, thus taking up a little of the electricity from that body, and bringing it near the gold-leaf electroscope, whereby the nature, and to some extent also the surface density, of the charge upon the electrified body may be determined.

If the proof-plane is touched to the surface of a charged hollow sphere, the distribution will be found to be uniform over the entire exterior suface; but if it is touched to the inside of the sphere, it will be found that there is absolutely no electricity whatever upon the interior. The charge, therefore, resides entirely upon the surface, and does not penetrate in the slightest degree to the interior, however highly electrified the conductor may be. A given quantity of electricity, accordingly, will charge a gilded pith-ball or a solid metal ball to the same degree, provided both are of the same size. Faraday constructed a room which he coated upon the outside with tin-foil, and thoroughly insulated from the ground. Upon entering this room he was able to perform the most delicate electrical experiments, although the outside was charged as highly as possible by electrical machines. The most sensitive electroscope within the room failed to show by the slightest indication the high electrification of the room upon the exterior.

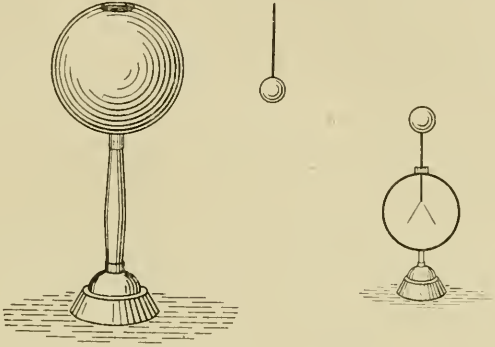

FIG. 194. The proof-plane.

The proof-plane may likewise be used to determine the distribution of electricity upon charged bodies which are not spherical. If the different portions of an electrified disc are examined in this way, Fig. 195, it will be found that the electrical density is much greater at the edges of the disc than on the more central portions of the surface. Similarly, the electrical density upon an egg-shaped conductor is found to be extremely

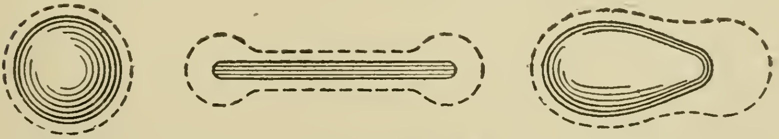

FIG. 195. The distribution of electricity upon a sphere, a disc, and an egg-shaped conductor.

irregular, being much greater upon the smaller end than upon the larger. Indeed, the surface density upon any part of a conductor, as will be shown later, is found to vary inversely as the radius of curvature of that part. The electrical density upon a point, therefore, would be infinitely great, inasmuch as the radius of curvature is infinitely small, were it not for the fact that a point upon or near an electrified conductor discharges that conductor, as will also be explained later. It is mainly upon this discharging power of points that the efficacy of the lightning rod depends.

309. Explanation of the Surface Charge. The explanation of the fact that, under ordinary circumstances, electrification is found only upon the exterior surface of a charged body is due to the repulsive forces which exist between similarly electrified bodies. Because of these forces, it is evident that, if a body is charged negatively, the electrons which are present in excess must repel one another; while, if a body is charged positively, the large numbers of atoms which have lost an electron, and which are therefore positively charged, must likewise repel one another. Accordingly, whether a body is charged negatively or positively, the elementary quantities of electricity constituting the charge must be constantly in a state of mutual repulsion, each forcing the others as far away as possible, that is, to the exterior surface of the body.

One exception only occurs to the general law that the electrification of a charged insulated conductor resides upon its outer surface. This is found when a charged body is lowered by an insulating rod or thread into the interior of a hollow insulated conductor, but without touching it. A charge is then found distributed not only upon the exterior surface of the conductor, but also upon the interior surface as well, as will be explained in a future paragraph.

310. The Action of a Charged Sphere Upon an External Point. In dealing with the action of one charged body upon another exterior to it, we shall assume for electrified bodies what Newton proved for gravitating masses of matter, namely, that a charged sphere acts upon an external point as if the entire electrification were concentrated at the center of the sphere. Indeed, Coulomb demonstrated experimentally the truth of this principle, in deducing the law of the force between two electrified bodies; and in the derivation of the unit quantity of electricity (p. 338), the distance r, between the two small spheres charged with quantities q and q', is taken without error to be the distance between their centers.

311. The Electric Intensity at a Point Infinitely Near a Charged Sphere. *By the electric intensity at a point is meant the number of dynes of force with which a small body charged positively with one unit of electricity will be acted upon when placed at that point.* Accordingly, the electric intensity infinitely near a charged sphere is the number of dynes of force with which the sphere will repel a unit of positive electricity when brought infinitely near to its surface.

In the preceding paragraph we have seen that, if a sphere is charged with electricity, its action upon exterior bodies is the same as if this charge were concentrated in a point at the center. Therefore, by the preceding paragraph, if one unit of positive electricity, that is, a small body charged positively with one unit of electricity, is brought infinitely near to the exterior surface of a sphere of radius R, charged with Q units of electricity, the unit is at a distance approximately equal to the radius R from the Q units with which the sphere is charged. The two will then repel each other with a force F, where

$$F = \frac{Q \cdot 1}{R^2}.$$

But

$$Q = 4 \pi R^2 \sigma,$$

and therefore

$$F = \frac{4 \pi R^2 \sigma}{R^2},$$

or

$$F = 4\,\pi\,\sigma.$$

The force, therefore, with which a charged sphere will repel a unit of positive electricity infinitely near its surface, that is, *the electric intensity infinitely near the surface of a charged sphere, is $4\pi\sigma$ dynes,* where σ is the surface density of the charge. It is to be observed that *the electric intensity is independent of the size of the sphere,* and is equal to $4\,\pi$ multiplied by the surface density of the charge.

As will be proved in the following paragraph, the electric intensity at all points throughout the interior of a charged conductor is zero. Inasmuch, then, as the electric intensity is zero upon the interior of a charged spherical conductor and $4\pi\sigma$ upon the exterior, the electric intensity changes by $4\pi\sigma$ in passing from the inside of a charged sphere to the outside. This proposition still holds true when the radius of the sphere becomes infinitely great, that is, when the surface of the sphere becomes a plane. From this it follows that the electric intensity changes by $4\pi\sigma$ in passing from one side of a charged plane to the other; or, in other words, if a plane is uniformly electrified on both surfaces, the electric intensity infinitely near this plane is $2\pi\sigma$ on each side. This law, which is thus found to hold true for a sphere and for a plane, is found also to apply to any surface whatever, hence the following general law: The electric intensity changes by $4\pi\sigma$ in passing from one side of any electrified surface to the other.

312. The Electric Intensity at a Point Within a Charged Conductor is Zero. In a previous paragraph (p. 149) it was shown that a hollow shell of gravitating matter exerts no resultant force of attraction upon a particle placed within. By a process of reasoning in every way similar to this, we may prove that a uniformly charged body exerts no force of attraction upon a unit of positive electricity placed upon its interior. In other words, it may be shown that the electric intensity at a point within a charged conductor is zero.

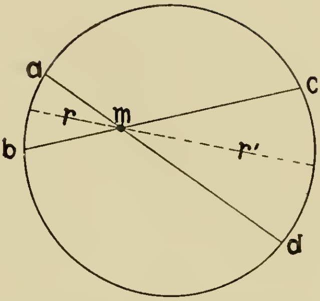

Fig. 196. A charged body exerts no force of attraction upon a unit of electricity placed within.

Let *a b c d*, Fig. 196, represent a closed conductor, a sphere for instance, charged positively and uniformly, and let *m* represent 1 positive unit of electricity placed at any point within. The unit at *m* may then be made the common vertex of two small similar cones, the bases of which, *a b* and *c d*, are at the distances *r* and *r′* respectively from the unit. The charge upon the base *a b* will then repel the unit at *m* with a force

$$\frac{ab \,.\, 1}{r^2},$$

acting towards the right, while that upon the base *c d* will repel the same unit with a force

$$\frac{cd \,.\, 1}{r'^2},$$

acting towards the left. The resultant of these two forces will be their difference, or

$$\frac{ab}{r^2} - \frac{cd}{r'},$$

Acting towards the right or the left according to their magnitudes.

By geometry, however, we know that the bases of similar cones are to each other as the squares of their altitudes, or that

$$\frac{ab}{r^2} = \frac{cd}{r'^2}.$$

But each member of this equation is identical with one of the forces acting on the unit at *m*, from which it is evident that these forces are equal in intensity and opposite in direction, or that the resultant force due to the portions *a b* and *c d* of the charged sphere acting upon this unit is zero.

In the same way the entire surface of the sphere may be divided into pairs of similar cones, for each of which the same holds true, so that the unit at *m* experiences no resultant force in any direction. The proposition is therefore proved that a uniformly charged conductor exerts no resultant force upon a unit of electricity placed within, or, in other words, the electric intensity at all points upon the interior of a charged conductor is zero.

313. The Flow of Electricity—The Electric Current. It has already been shown (p. 343) that the distinction between a conductor and a non-conductor, or insulator, lies in the fact that the conductor contains among its molecules large numbers of swiftly moving atoms or particles of negative electricity, called the free electrons, while the non-conductor contains relatively few, or practically none, of these minute bodies. Since the free electrons in a conductor are thought to be in thermal equilibrium with the molecules, that is, to bombard the molecules and to be in turn bombarded by them, their motions, though high in speed, take place in all possible directions, and, therefore, produce no resultant effect. When, by any means, these free electrons are compelled to move in a particular direction, electricity is thereby transferred from one point to another, and a current of electricity is said to flow. The flow of the electric current is thus thought to be simply the orderly, or directed, movement of the free electrons in the conductor.

Nothing would be more natural than to suppose that the direction in which the current flows would be the same as that in which electricity is thus transported. Such, indeed, was confidently believed to be the case as long as the one-fluid theory of Franklin (p. 340) was accepted. By this hypothesis, a positively charged body was thought to contain an abnormal amount of electricity, which would flow to the earth from the body when the latter was touched. The direction of the current was therefore interpreted to be the direction in which this discharge took place, that is the direction in which the positive charge upon a body flowed to the earth. Thus, the positive direction of the current was assumed to be the direction of the motion of the positive charge; and this is the direction which, through many years of constant use, is now universally understood to be the direction in which the current flows. It is true, as we now believe, that it is not the positive charge that flows, but that it is the negative electrons that move, and that these move in the direction opposite to that in which the positive charge would move if it were free to do so.

A moment's reflection, however, will show that these two conceptions are identical in import. By analogy with the principle in speech of the double negative implying an affirmative, we can see that the negative electrons moving in a negative direction produce the same result as the positive charge moving in the positive direction. Accordingly, the effect of the negative electrons moving in the negative directions is the same *as if* the positive charge moved in the positive direction, or in the direction in which the current is ordinarily assumed to flow. If Franklin had given the opposite signs to the electric charges, that is, if he had called the resinous electrification positive, and the vitreous negative, the electrons, as we now understand them, would have been positive atoms of electricity, and the direction in which they flow would likewise have been positive. The direction of the current would then have been that of the motion of the electrons, and the conception of the flow of the current would have been greatly simplified.

In the work which follows, therefore, we shall endeavor to determine and trace the movements of the electrons, which indicate the actual processes taking place within the conductor. We can not, however, speak of this direction as the direction of the current, but are constrained to conform to popular usage, and to say that the current flows in a direction opposite to that in which the electrons move.

CHAPTER XXIII.

THEORY OF POTENTIAL.

314. Difference of Potential. In dealing with negatively and positively charged bodies, we must keep constantly in mind the idea that a negatively charged body is one which has upon it an abnormal number of electrons, while a positively charged body is one which has lost some of the electrons that were upon it in its natural state. Accordingly, if a negatively charged body is connected by a conductor with the earth, this excess of electrons will flow from the body to the earth, leaving the body once more in its neutral state. In the same way, if a positively charged body is thus connected with the earth, electrons will flow from the earth to the body, thereby restoring to the latter its natural number of these particles. In either case, as soon as the transport of electrons has taken place, the body loses all traces of electrification, or becomes, as we say, "discharged."

When a neutral body is connected with one positively charged, electrons will pass from the former to the latter, leaving the former positively charged, and the latter also positively charged, but less highly than before. When a neutral body is connected with one negatively charged, the electrons upon the latter will be distributed over the surface of both bodies, and both will be charged negatively. When two bodies, one charged positively and the other negatively, are connected together, the electrons upon the negatively charged body will be divided between the two, and both will be positively or both will be negatively charged, according to the number of electrons that the negatively charged body may have originally contained.

In these, and in all other instances of the directed motion of the electrons, we may attribute this movement, not merely to the mutual attractions or repulsions of electrified particles, but to a difference in the electrical condition of the bodies, known as a difference of potential. *Difference of potential may be defined as a difference of electrical condition, due to which elec-*

trons will flow from one body to another, or would flow if the two were connected by a conductor. The region towards which the electrons move is said to be a region of high potential, while that from which they come is, in the same way, called a region of low potential. A positively charged body, therefore, if uninfluenced by other electric charges, is a region of high potential, while, for the same reason, a negatively charged body is a region of low potential. It is thus seen that when a current of electricity flows, the electrons move from the region of low potential towards the region of high potential, although, out of deference to the universal usage, we are obliged to say that the *current* flows from the region of high potential towards the region of low potential.

In determining the movement of the electrons, and therefore in ascertaining the condition of the potential, as to whether this is high or low, a comparison is made with the earth, which is always taken as the zero of potential. Accordingly, a body is at high potential if, when it is connected with the earth, electrons flow from the earth to that body, and it is at low potential if, when it is thus connected, electrons flow from the body to the earth. It must be remembered that this is the only criterion which enables us to tell whether the potential of a body is high or low, for, if other electric charges influence the body, the potential of the body cannot be determined from its charge, nor, indeed, can the charge of a body be determined from its potential. In a single instance only is it possible to predict the potential of a body from its charge, and this instance occurs when the body is standing entirely alone and uninfluenced by any other electrified body. Under these conditions the conclusion may be drawn that the body is at high potential if it is charged positively, that is, if it has upon it less than its normal number of electrons, and that it is at low potential if it is charged negatively, or has upon it an excess of electrons above its normal number. In general, however, *the character of the potential of a body may be determined only by connecting that body with the earth.* If the flow of electrons takes place *from the earth to the body,* the body is at *high* potential, but if the flow takes place *from the body to the earth,* the body is at *low* potential.

315. The Theory of Potential. Let us suppose that we have an insulated conductor charged positively and removed from all other electrified bodies, Fig. 197, and let us suppose that we bring into its neighborhood a unit of positive electricity, that is, a small body, a pith ball, for instance, charged positively with one unit of electricity. If there are Q units upon the charged body, we know by Coulomb's law (p. 338) that the unit of positive electricity will be repelled by the charged body with a force of F dynes, where

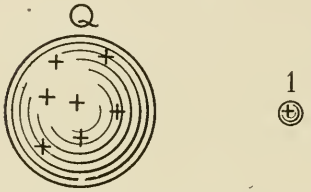

FIG. 197. A positively-charged body repels a unit of positive electricity placed near it.

$$F = \frac{Q \cdot 1}{r^2},$$

r being the distance between them. If the unit of positive electricity is brought up nearer to Q, the force of repulsion will be greater, and if it is removed farther from Q, the force of repulsion will be less.

If the unit of positive electricity is prevented at any point from moving away, a force equal to the repellant force at that point must be exerted to prevent it from moving. Moreover, if this unit is forced to move nearer to Q against the repellant force, *work* must be done upon it equal to the product of the force thus exerted and the space through which it is moved. When the unit is at infinity, the distance between the two is infinitely great, that is,

$$r = \infty,$$

and

$$F = 0,$$

which indicates that at infinity the unit is acted upon by no force of repulsion from Q. Whenever r is finite, the force exerted is finite, and this force, although it is small when r is very great, becomes greater the nearer the unit approaches the charged body.

If the unit is supposed to be moved up from infinity toward the charge Q, Fig. 198, a perfectly definite amount of work will have to be expended against the repellant force due to Q in bringing it up to the point B. If the unit is forced to move still nearer to Q as far as the point A, a still greater amount of work will have to be performed. When the unit is at the point B, a certain number of ergs of work have been performed upon it in bringing it from infinity up to this point, and the potential energy which it now possesses is measured by the work that has been done.

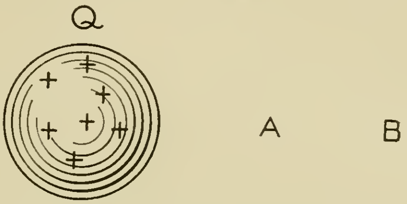

FIG. 198. More work must be expended in bringing a unit of electricity from infinity to the point A than to the point B.

There is, therefore, an attribute of every point about a charged body of such nature that a unit of positive electricity when brought to that point will possess potential energy. The potential energy is not developed until the unit of positive electricity is there; but *the peculiar condition of that point, due to the presence of the charged body, exists before the unit of electricity is brought up*, and is known as the *potential* of that point. The electrical potential at a point, therefore, is that condition of the region about a charged body in virtue of which a unit of positive electricity when brought to that point will possess potential energy; and *the measure of the electrical potential at a point is the number of ergs of potential energy which the unit of positive electricity will possess when brought to that point.*

Accordingly, the electrical potential at the point B, Fig. 198, is the number of ergs of potential energy which the unit of positive electricity will possess when brought to B. We can see that this is so, because if the unit of positive electricity were brought to B and released, it would move off to infinity, being repelled by Q, and in so doing it would be able to perform just as many ergs of work as were expended upon it in bringing it from infinity up to B. The attribute of B, therefore, before the unit is brought up, is such that when the unit of positive electricity is brought from infinity to that point, this unit will posses an amount of potential energy equivalent to the number of ergs of work which have thus been expended. Accordingly, *the electrostatic potential at a point is the number of ergs of work that would be required to bring a unit of positive electricity from infinity to that point.*

More work would be required to bring up the unit of positive electricity from infinity to the point A than from infinity to the point B, and therefore the point A is at higher potential than the point B. From this it is seen that the difference of potential between A and B is the number of ergs of work that are required to carry a unit of positive electricity from B to A. In general, *the difference of potential between any two points is the number of ergs of work that must be expended in carrying a unit of positive electricity from one point to the other.* If only one erg of work is expended in bringing the unit of positive electricity from one point to the other, these points have a difference of potential of unity between them. Accordingly, *two points have unit difference of potential between them when one erg of work is required to bring the unit of positive electricity from one point to the other.*

The expressions for the potential at a point and for the difference of potential between two points are of the greatest importance and are derived in the following paragraph.

316. Measurement of Potential. As before, let Q, Fig. 199, represent a body charged positively with Q units of electricity, and let A and B be two points in a straight line from Q and at distances r and r' respectively from it. If a unit of positive electricity is brought up to B, this unit will be repelled by Q with a force

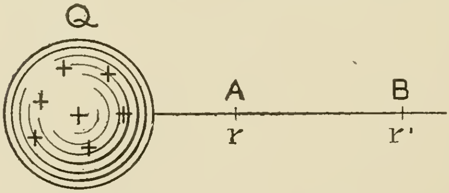

FIG. 199. The measurement of the difference of potential between A and B.

$$\frac{Q\,.\,1}{r'^2},$$

and if this unit is brought up to A it will be repelled with the force

$$\frac{Q\,.\,1}{r^2}.$$

In order to measure the difference of potential between A and B, it is necessary to determine the work which must be done in carrying a unit of positive electricity from B to A. This work is equal to the product of the force with which this unit is repelled and the distance

$$r' - r$$

through which this force must be exerted between these two points. The force by which this distance is to be multiplied cannot be the force at A, because this is too large, nor can it be the force at B, for this is too small. We may, therefore, speak only of an *average* force between A and B; and the work which must be done in carrying the unit of electricity from B to A is the product of this average force and the distance $r' - r$. Moreover, this average force is the mean of the forces at A and B, since the force varies continuously and according to the same law throughout this distance. In obtaining this average, we have to deal with forces which vary inversely as the *second* power of the distance, and hence it is necessary that this mean should be the geometrical mean, which is the square root of the product of these forces, and not the arithmetical mean, which is half their sum. Accordingly, since the force at A is

$$\frac{Q}{r^2},$$

and the force at B is

$$\frac{Q}{r'^2},$$

the average force between A and B is

$$\sqrt{\frac{Q}{r^2} \cdot \frac{Q}{r'^2}},$$

that is

$$\sqrt{\frac{Q^2}{r^2 \,.\, r'^2}},$$

or

$$\frac{Q}{r\,r'}.$$

Since the distance between A and B is

$$r' - r,$$

the work done in carrying the unit from B to A is

$$\frac{Q}{r\,r'}\,(r' - r),$$

or

$$Q\left(\frac{1}{r} - \frac{1}{r'}\right).$$

Accordingly, the difference of potential between A and B is

$$Q\left(\frac{1}{r} - \frac{1}{r'}\right).$$

If V represents potential, then V_A represents the potential at A and V_B represents the potential at B, and the difference of potential between A and B is represented by

$$V_A - V_B = Q\left(\frac{1}{r} - \frac{1}{r}\right).$$

If B is taken at infinity, then

$$r' = \infty,$$

and

$$\frac{1}{r'} = 0.$$

In this case likewise

$$V_B = 0,$$

and the difference of potential between A and infinity, that is, the absolute potential at A is

$$V_A = \frac{Q}{r}.$$

It is to be noted that

$$\text{The } \textit{force} \text{ at } A = \frac{Q}{r^2},$$

and that

$$\text{The } \textit{potential} \text{ at } A = \frac{Q}{r}.$$

Thus it is seen that while the force at a point due to the presence of a charged body varies inversely as the square of the distance from that body, the potential at the same point due to the same charged body varies inversely as the first power of the distance.

317. Equipotential Surfaces. If we had gone out from the charged body in any other direction, we should have been able to find another point at which the potential is the same as that at the point A. There exists, therefore, about a charged body an infinite number of points, at any one of which the potential is the same as at any other one. Through all such points it is, accordingly, possible to lay an imaginary surface such that every part of that surface shall be at the same potential as every other part. Such a surface is known as an *equipotential* surface, because the potential over its entire extent is constant. The equipotential surface about a charged metallic sphere removed from all disturbing influences is itself a sphere concentric with the conductor; but if the conductor is of any shape other than spherical, or if other charged bodies are anywhere in the neighborhood, the equipotential surface is not a sphere. Whatever the shape of the charged conductor may be, it itself is always an equipotential surface, because if the potential were not the same on every part, electrons would flow from one part to another until the whole were at one potential. The surface of every charged conductor, therefore, however irregular this may be, is an equipotential surface.

Inasmuch as the potential due to a charged body varies inversely as the distance from that body, it is possible at another distance to find another series of points, for all of which the potential has a uniform value. Through these points we may likewise draw another equipotential surface, such that over its entire area the potential is everywhere constant, though of different value from the first. In a similar way a third and a fourth and many other equipotential surfaces may be drawn, and indeed the entire region about a charged body may thus be divided into a series of equipotential surfaces. If these surfaces are so taken that one erg of work is required to carry the unit of positive electricity from one surface to the next, the potential of each surface differs by unity from that of the succeeding one. These equipotential surfaces between which there exists unit difference of potential are situated relatively near together at short distances from the charged body and become more and more widely separated from one another as the distance from the charged body increases. This is due to the fact that the force becomes rapidly less as the distance from the charged body increases, and therefore the distance through which this force must be moved in order that one erg of work may be done becomes increasingly great.

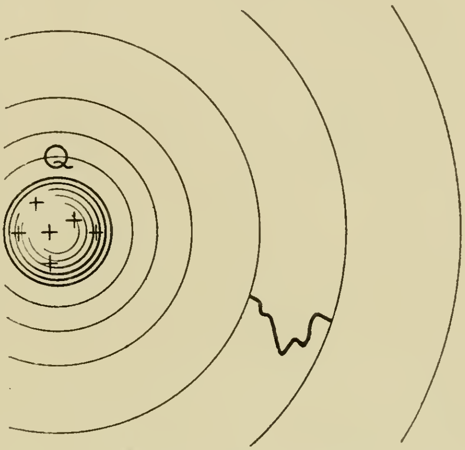

FIG. 200. Equipotential surfaces.

In this connection it is seen that no work is required to carry the unit of positive electricity from one point upon an equipotential surface to any other point on the same surface. This is true because the potential over this entire surface is constant, and work can be done only when electricity is carried from one potential to another. Therefore, if the positive unit is carried from one equipotential surface to another, *the work done is independent of the path* which is followed, and is equal to the average force multiplied by the perpendicular distance between the two surfaces. However, devious a path this unit may follow as it passes from one equipotential surface to another, Fig. 200, its motion may always be resolved into two components, one of which is parallel to the equipotential surface, and the other of which is perpendicular to this surface. It requires, however, no work to carry a charge parallel to an equipotenial surface, hence the entire work is expended in carrying this unit along the normal component of the path, that it, along the perpendicular line between the two equipotential surfaces.

318. Electrostatic Lines of Force. The region all about a charged body, wherever the influence or presence of the charged body may be detected, is known as an elctrostatic field of force, or simply an electrostatic field, Fig. 201. If a unit of positive electricity were placed in such an electrostatic field, it would tend to move from the region of high potential where it is placed toward a region of lower potential, and in so moving it would describe a certain path. Since it always moves under the action of the existing electrical forces, its motion at each point of its path indicates the direction of the resultant force at that point, and the path which this moving unit of positive electricity would thus trace out is known as a *line of force.* A line of force in an electrostatic field may therefore be defined as the path along which a unit of positive electricity would move if it were placed in this field, or it may be defined as a line so drawn in an electrostatic field of force that every point of the line is tangent to the direction of the resultant force at that point.

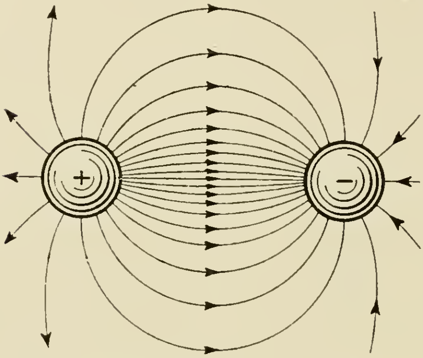

FIG. 201. Electrostatic lines of force.

Since a unit of positive electricity always tends of itself to move from a region of high potential toward a region of low potential, a line of force in general has its origin upon a positively charged body and ends upon a negatively charged body. Even if a positively charged body stands apparently all alone by itself, the lines of force go out from it and end upon the walls and ceiling of the room, where, as we shall learn later, a negative charge exists by induction, equal in amount to the positive charge upon the body.

If we mark out any small area upon the positively charged body and construct lines of force around the periphery of this area, we shall have a tube of force, Fig. 202, the walls of which consist of lines of force. The remote end of this tube of force will intercept a certain area upon the negatively charged body, and will include within this area a negative charge equal in amount

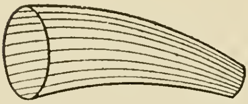

FIG. 202. A tube of force.

to the positive charge upon the area at the origin of the tube. If the charge at each end of such a tube is unity, the tube of force is called the unit tube.

319. Lines of Force and Equipotential Surfaces are Perpendicular to Each Other. Inasmuch as no component of the force exists parallel to an equipotential surface, it is evident that a line of force must go out normally from one equipotential surface and must meet the next one normally, or, in other words, lines of force and equipotential surfaces are everywhere perpendicular to each other. Accordingly, if the equipotential surfaces about a charged sphere form a system of concentric spheres, the lines of force are everywhere radii of these spheres, Fig. 203. If the equipotential surfaces are surfaces of irregular shape, as is the case when several differently charged bodies are near together, the lines of force nevertheless form a system of lines proceeding from regions of high potential and extending to regions of low potential, and everywhere meet the equipotential surfaces perpendicularly.

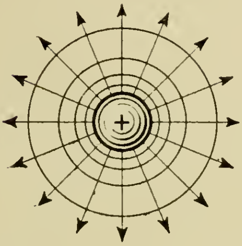

FIG. 203. Lines of force and equipotential surfaces about a positively-charged sphere.

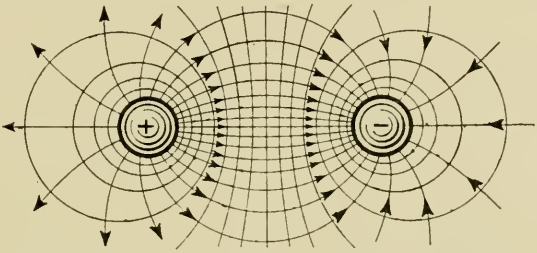

FIG. 204. Lines of force and equipotential surfaces about two spheres, one charged positively and the other negatively.

320. The Theory of Electrostatic Lines of Force. While we really know nothing about the mechanism by which a force is exerted at a distance from an electrified body, and therefore while we are unacquainted with the real cause of the flow of electricity, as well as the reason for the attraction and repulsion of electrified bodies, it is nevertheless believed that the ether all about a charged body is thrown into a state of strain. The ether may be conceived to be a jelly-like, highly elastic substance, which is compressed, or distended, or in some way deformed, as soon as a body is electrified. It is probably because of these distortions in the ether that a flow of electricity is produced, or that the mechanical attractions and repulsions of electrified bodies take place. Accordingly, when two similarly electrified bodies are brought near each other, the ether in the region between them may be considered as compressed, and this compression of the ether causes the two bodies to move apart. In the same way, when a positively charged body is brought into the neighborhood of a negatively charged body, the ether in the intervening space is regarded as distended or stretched, and the two bodies are accordingly pulled toward each other. Moreover, the lines of force are everywhere tangent to the direction of the force at every point throughout the field, hence it is undoubtedly true that the lines of force indicate the directions along which the ether is strained in the neighborhood of an electrified body.

These strains in the ether produce a phenomenon precisely analogous to tightly stretched elastic cords, and it is thought that it is because of the tendency of these lines to contract that the attraction is produced between a positively and a negatively charged body. In the same way, similarly electrified bodies repel each other, because in the space between two such bodies a number of these elastic, rubber-like lines of force are crowded, and this forces the two bodies apart. Moreover, lines of force never intersect, because if they were to do so, the resultant force would have two directions at the point where the intersection occurred, which is impossible.

When a glass rod is electrified by being rubbed with a piece of silk, the electrons normal to the two are redistributed, the glass losing some of the number which it originally had, while the silk acquires those which the glass loses. Strains in the ether extend therefore from the positively charged, high potential glass rod to the negatively charged low potential silk, and these tend to draw the two together. Accordingly more work is required to separate the glass rod from the silk after they are rubbed than before they were electrified; and the energy which these, electrified bodies possess, therefore, is this additional work that has been done in separating them. Thus it is seen that electricity itself is not a form of energy, but that the energy resides in the strained condition of the ether all about the charged body.

321. The Potential of a Charged Sphere. When a point is situated at the distance r from a spherical conductor charged with Q units of electricity, its potential, as we have seen (p. 359) is

$$V = \frac{Q}{r}.$$

Moreover, this conductor, as has likewise been seen, (p. 346), acts upon exterior bodies as if its entire charge were concentrated in a point at its center.

If we consider a point exterior to a charged sphere of radius R, but infinitely near to it, this point is then at the distance R from the center, and its potential, and therefore that of the entire sphere, is

$$V = \frac{Q}{R},$$

since the surface of the sphere is the limiting case of an external point. Furthermore, the potential is constant, not only over the surface of the sphere, but also over the surface of any charged conductor in equilibrium, however irregular its distribution of charge may be, since otherwise electrons would flow until there were no longer a difference of potential. The surface of any charged conductor is therefore an equipotential surface.

It may also be shown that the potential is constant throughout the interior of a charged conductor, and, indeed, is the same in value as that of the conductor itself. This may be seen at once by considering that the resultant force acting upon a unit of electricity within a charged conductor is everywhere zero (p. 348), so that no work would be required to move the unit from one part of the interior to another. Accordingly, the potential is constant throughout, and equal to that of the conductor itself, since the inner surface of the conductor is the limiting position of an interior point. Only in the case that the conductor surrounds an electric charge, that is, that an electric charge is brought within the conductor without touching it, is there electrification and also a difference of potential within the conductor. Then and only then can lines of force exist within a charged conductor, for lines of force must extend from a region of high potential to a region of lower potential.

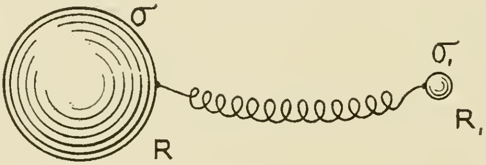

FIG. 205. The surface density varies inversely as the radius of curvature.

322. The Surface Density Upon a Charged Body Varies Inversely as the Radius of Curvature. If two charged spheres of radii R and R' are connected by a conductor, Fig. 205, the electrification will be so distributed between the two that all parts of the system will be at the same potential, as was shown in the preceding paragraph. If it is then found that one sphere is charged with Q units and the other with Q' units, the potential of the first will be

$$\frac{Q}{R},$$

while the potential of the second will be

$$\frac{Q'}{R'}.$$

Since there can be no difference of potential between two bodies connected by a conductor, it must follow that the potentials of these spheres are the same, and that

$$\frac{Q}{R} = \frac{Q'}{R'}.$$

But (p. 344), the quantity of electricity with which a sphere is charged is

$$Q = 4 \pi R^2 \sigma,$$

where R is the radius and σ is the surface density. Accordingly,

$$\frac{4 \pi R^2 \sigma}{R} = \frac{4 \pi R'^2 \sigma'}{R'},$$

or

$$R \sigma = R' \sigma',$$

from which

$$\frac{\sigma}{\sigma'} = \frac{R'}{R}.$$

It is thus seen that if two spheres are connected by a conductor and charged, the charge will be so distributed that both will have the same potential, after which the surface densities on the two spheres are inversely as their radii. If the two spheres are now conceived simply as parts of an extended conductor having different radii of curvature, the potential will be everywhere the same, and the surface densities upon the various parts will then be inversely as the radii of curvature of these parts. This explains the unequal distribution of the charge upon conductors which are not spherical in form.

323. The Discharging Power of Points. If we now conceive one of the above spheres to grow smaller, its surface density will become greater in the same proportion as its radius becomes less; and when its radius is indefinitely small, its surface density would be indefinitely great, were it not for the fact that the diminishing sphere would then become a point, and a point, as we have seen (p. 346), serves to discharge a conductor.

The underlying cause for the discharging power of points is found in the rapid fall of potential which occurs in their neighborhood. This may at once be seen by constructing the equipotential surfaces in the neighborhood of a body provided with a sharp point. Thus, if *P*, Fig. 206, is a metallic point on the charged conductor *AB*, the potential at *P* is the same as that upon the conductor itself, since the potential of a conductor must be constant throughout. Moreover, if *C* is a conductor in contact with the earth, and therefore at zero potential, just as great a fall of potential must exist between the point *P* and *C* as exists between the conductor *AB* and *C*. Accordingly, the equipotential surfaces are crowded close together in the vicinity of *P*, and the fall of potential gradient is here very rapid.

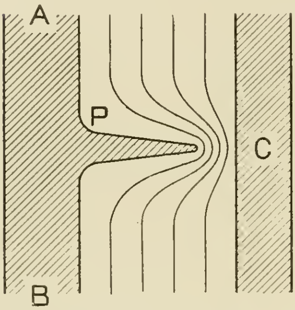

FIG. 206. The rapid fall of potential near a point, producing ionization.

A chance electron which may be present in the air, being attracted or repelled by the point, then passes through a great difference of potential, and thus acquires an enormous velocity. So great is its energy, because of its increased velocity, that it disintegrates an air molecule upon which it impinges, disengaging electrons from its atoms, thus leaving the latter positively charged. These newly-released particles produce similar results, so that the region about the point is quickly filled with a swarm of these minute charged bodies, which then act as carriers of electricity to or from the point, thereby discharging it.

Air which is thus filled with electrons and with positively charged atoms that have lost an electron, becomes at once a conductor of electricity, and is said to be *ionized*.

Toirel

CHAPTER XXIV.

ELECTROSTATIC INDUCTION.

324. Electrostatic Induction. If a positively charged body *A*, Fig. 207, is brought near a long, insulated, neutral conductor *B*, but not so near that electricity can pass from one to the other, the free electrons in *B* are redistributed, the end near *A* becoming negatively electrified, and that remote from *A* positively electrified. If communication is now established between any part of *B* and the earth, the upper end loses all traces of electrification, while the lower end becomes still more strongly electrified. If the connection with the earth is now broken, and the body *A* is removed or discharged, the entire surface of *B* will be found to be negatively electrified.

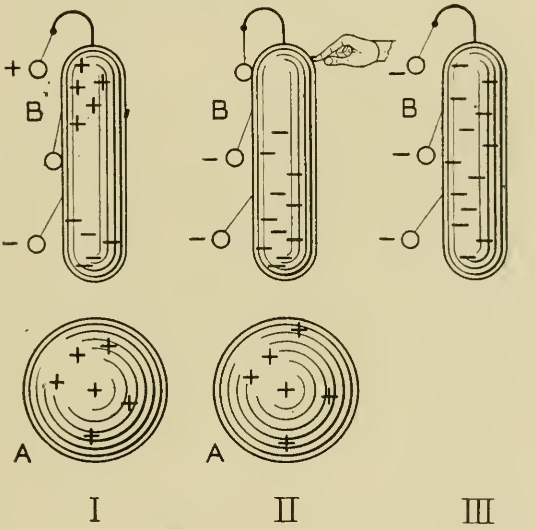

Fig. 207. The various stages of electrostatic induction.

The explanation of the electrification of *B* in the three cases may be given as follows, first, from the point of view of the mutual action of electrified particles.

The positive charge upon *A* attracts the free electrons upon *B*, causing them to flow from the upper end towards the lower, thus leaving the upper end charged positively, and charging the lower end negatively. This produces the electrification of *B* in the first case. When *B* is connected by a conductor with the earth, electrons flow from the earth into *B* and distribute themselves uniformly over it, thus making the upper part neutral, and electrifying the lower part more strongly negatively. This, then, represents the electrification of *B* in the second case. When *A* is now removed or discharged, the electrons are no longer attracted in excessive numbers towards the lower part of *B,* but distribute themselves over the entire conductor according to the law of the surface density for a conductor of this form; and as electrons had flowed from the earth into *B* when the latter was connected with the earth, this conductor is now electrified negatively throughout. This constitutes the electrification in the third case.

Also from the point of view of the theory of potential we may explain the electrification of *B,* this method, however, being in every way equivalent to the first. In doing this, it is necessary to consider the potential of *B* in each of the three cases. In the first case, it is to be observed that, except during the exceedingly short time required for the rearrangement of the electrons, there can be no difference of potential between the different parts of *B,* inasmuch as *B* is a conductor. Notwithstanding the fact that one end of *B* is charged positively and the other negatively, *both ends are at the same potential.* In the third case, when the body *A* is removed, *B* is found to be charged negatively throughout. Since *B* had its normal number of electrons and was at zero potential before it was brought into the presence of *A,* electrons must have flowed from the earth into *B* in the second case, when *B* was connected with the earth, or grounded. It is evident, therefore, that *B* must have been at high potential before communication between it and the earth was established, and this could have been brought about only by the presence of the body *A.* Accordingly, the simple presence of the body *A,* which itself is at high potential, raises the potential of the body *B.*

In general, the presence of a body of high potential raises the potential of all other bodies in its neighborhood, and, conversely, the presence of a body of low potential lowers the potential of all other bodies in its neighborhood. The nearer a body is brought to a region of high potential, the higher does its own potential become.

It is now not difficult to understand how, in the first case, the free electrons normal to *B* are rearranged, and the two ends of this body become oppositely charged. When *B* is brought into the presence of *A*, the lower end of *B* is brought nearer to the region of high potential than the upper end, and, for an instant, accordingly, the lower end is at higher potential than the upper end. The free electrons in *B* are therefore redistributed, flowing from the upper end to the lower, leaving the upper end charged positively, and charging the lower end negatively, this flow continuing until there is no longer a difference of potential between the two ends.

When the electrons are at rest upon a conductor or a system of conductors, it is evidence that there can be no difference of potential between any parts of the system. Accordingly, in the first case, after the redistribution of the electrons has been effected, there is no difference of potential between the two ends of *B*, although one end is now charged negatively and the other positively. Moreover, both ends are now not only at the same potential, but they are at high potential, since both are in the presence of *A*. Therefore, since the potential of *B* is constant throughout, electrons will flow from the earth in equally large numbers to both ends of *B* whenever any part of *B* is connected with the earth, and will continue to flow until *B* is at zero potential.

When *A* is removed, the space which *A* occupied is lowered in potential, and this lowers the potential of *B* . *B* is, accordingly, negatively charged, since electrons flowed from the earth upon it; and it is also at low potential, inasmuch as its potential was zero when it was connected with the earth, and this has now been lowered, due to the removal of the high potential *A*. Therefore, when *B* is once more connected with the earth, the excess of electrons which it now contains flows to the earth, raising its potential again to zero, and restoring this body to its normal, neutral condition.

The redistribution of electrons upon *B* due to the presence of *A* is known as induction, and the body *B* is said to be charged by induction due to the influence of the body *A*. In general, a body of high or low potential induces charges upon all other bodies in its neighborhood.

It is interesting to observe the relation existing between the lines of force (p. 362) which radiate from a body of high potential and the charges which this body induces upon all conductors in its electrostatic field of force. The lines of force which emanate from *A* and are intercepted by *B*, Fig. 208, fall upon the end of *B* which is negatively electrified, while lines of force connect the remote, positively electrified end of *B* with other regions which are still lower in potential. It is thus seen that lines of force, in connecting regions of high potential with those of lower potential, really emanate from a positively charged body and end upon a body, or a portion of a body, which is negatively charged. The fact that lines of force enter the end of *B* near *A* and leave the end of *B* remote from *A* indicates that the end of *B* near *A* is charged negatively, and that the end of *B* remote from *A* is charged positively. Any body or portion of a body, therefore, which has electrostatic lines of force entering it is charged negatively, and any body or portion of a body which has electrostatic lines of force emerging from it is charged positively. It would be impossible for lines of force to connect two bodies similarly and equally charged.

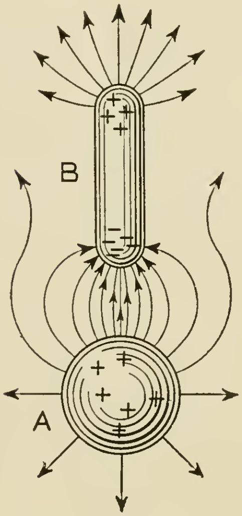

FIG. 208. Lines of force in electrostatic induction.

An electrostatic line of force, accordingly, always originates in a positively charged body or the positively charged portion of a body. Even if a positively charged body is apparently situated entirely alone and uninfluenced by other electrified bodies, the lines of force which emanate from it end upon the ceiling, the walls, and the floor of the room, and here develop negative charges by induction equal in magnitude to the original positive charge upon the electrified body. If a neutral body has lost electrons so that it is positively charged, the body itself is then the origin of a number of lines of force which emanate from this body and end upon a negatively charged body somewhere in the neighborhood. In a similar way, if electrons have been added to a body so that the latter is charged negatively, this body is the ending place of lines of force which come to it from surrounding positively charged bodies. If, however, the normal electrons upon a body have been simply rearranged by induction, as is the case with the body *B*, Fig. 208, so that one end is charged positively and the other negatively, the body itself is not the origin of any lines of force, for just as many lines of force leave the body from the positively charged end as enter it from the negatively charged end.

The various stages of this experiment illustrate the important principle that when charges of electricity in the neighborhood act inductively upon a given body, the nature of the charge upon the body cannot be determined from its potential, nor can the potential of the body be determined from its charge. Only when a body stands alone and is wholly uninfluenced by other electrified bodies can it be assumed that the body is at high potential if it is charged positively, and at low potential if it is charged negatively. In general, however, the charge does not depend upon the potential, nor does the potential depend upon the charge.

325. The Gold-Leaf Electroscope Charged by Induction. In electrifying the leaves of the gold-leaf electroscope, if the knob is actually touched by a positively electrified glass rod it will be found that the high potential of the rod will agitate the leaves so violently that there is danger of tearing or destroying them. It is better, therefore, always to electrify the leaves by induction, after the manner in which the body *B* (p. 369) was electrified. This process may likewise be explained either by the attraction and repulsion of electrified particles, or by the influence which one electrified body exerts upon another, that is, by the theory of potential.

If it is desired to charge the electroscope positively, we first bring near the instrument a negatively charged body, such as an electrified rubber rod, Fig. 209. This will repel to the gold-leaves the free electrons which are present in the knob and connecting rod, leaving the knob charged positively, and charging the leaves negatively. This is the first case in the analogy with the bodies *A* and *B*. If the knob or any part of the connecting rod is now touched by the hand, electrons will flow from every part of the system, leaving the leaves entirely neutral, and charging the knob still more strongly positively, thus forming the second case. If the connection with the earth is now broken and the rubber rod then removed, the knob, the rod, and the gold-leaves will be found to be positively charged, since electrons flowed away from this system when it was connected with the earth. This constitutes the third case.

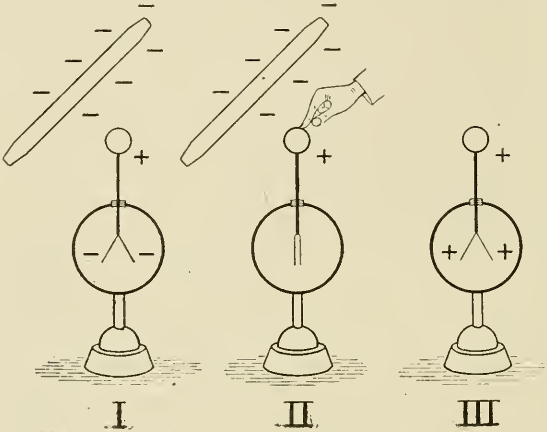

FIG. 209. The gold-leaf electroscope charged by induction.

By the theory of potential, this process is quite as simply explained, and is, perhaps, even more illuminating. Here the negatively charged rod takes the place of the body *A*, while the electroscope itself represents the body *B*. The low potential of the negatively charged rod lowers the potential of all bodies in its neighborhood, and hence lowers the potential of the electroscope. Moreover, the knob of the electroscope is nearer the electrified rubber rod than are the gold-leaves, therefore, for an instant, the knob is at a lower potential than the gold-leaves, that is, for an instant, the leaves are at a higher potential than the knob, and, as a consequence, electrons flow from the knob to the leaves. The electrons normal to the electroscope are therefore rearranged, the gold-leaves becoming negatively, and the knob positively, electrified, but the leaves and the knob are now at the same potential. This forms the first case.

If, while the rubber rod is still in place, the low potential knob is now connected with the earth, electrons flow from all parts of the electroscope to the earth, leaving the leaves unelectrified, charging the knob more strongly positively than before, and raising the potential of the entire system to zero. This corresponds to the second case. If the connection with the earth is broken and the electrified rod then removed, the space which the rod occupied becomes at higher potential, and this raises the potential of the electroscope, which thus becomes at high potential. Moreover, when connection with the earth was established, electrons flowed from the electroscope to the earth, leaving a deficit of electrons upon the electroscope, which means that the latter is now charged positively. This is the third case. By induction, therefore, the potential of the electroscope was lowered through the low potential of the rubber rod, so that, when connection was made with the earth, electrons flowed from the low-potential knob and leaves to the earth, leaving the system charged positively.

326. The Electrophorus. The electrophorus is a device for obtaining electricity in unlimited quantities by induction. This apparatus consists of a plate of hard rubber or resin, and a metal disc, or cover, provided with an insulating handle. The hard rubber plate is electrified negatively by rubbing or beating with cat's fur, and the cover is then lowered until it rests upon the plate. The cover is then touched, or otherwise connected with the earth, after which, when it is lifted by its insulating handle, it is found to be positively electrified. As in all other instances of static induction, we may look upon the charging of the cover of the electrophorus from two different points of view, which, however, are indentical in their significance.

From the view that similarly electrified bodies repel one another, we see that when the cover, held by its insulating handle, is brought near the negatively charged rubber plate, the electrons, which are in excess upon the plate, repel the free electrons of the metal cover, thus charging the lower surface of the cover positively, Fig. 210. Indeed, the cover is brought down until it rests upon the hard rubber plate, without, however, discharging it, since the inequalities of both surfaces permit of contact in but three or four points, which does not materially affect the original electrification of the plate. If contact is now made between the cover and the earth, electrons flow from both surfaces of the cover to the earth, leaving the lower surface more highly positively charged than before, and the upper surface neutral. When the connection with the earth is now broken and the cover lifted, the latter is found to be charged positively, since electrons escaped during contact with the earth.

From the view that an electrified body influences the potential of the region all about itself, we can see that the negatively electrified rubber plate, representing the body *A* (p. 369), is at low potential itself, and therefore lowers the potential of all neighboring bodies. Accordingly, when the cover, which corresponds to the body *B*, is brought down upon the plate, it is brought into a region of low potential, and so becomes itself at low potential. Moreover, the lower surface of the cover is nearer the region of low potential than is the upper surface, and is therefore at lower potential. For an instant, then, there is a difference of potential between these two surfaces, which causes the free electrons to flow from its lower surface to the upper, leaving the former charged positively, and charging the latter negatively. Though oppositely electrified, both surfaces are now at the same potential, and this is low, since both are in the region of the negatively electrified, low potential hard rubber

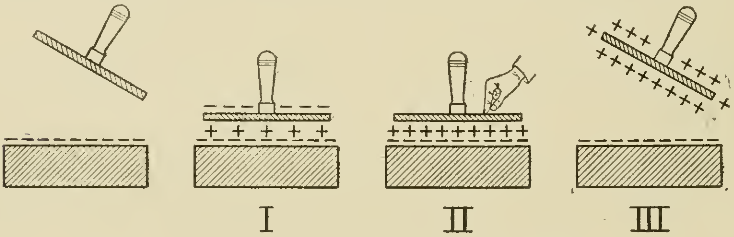

FIG. 210. The action of the electrophorus.

plate. This is the first case in the analogy with the bodies *A* and *B*.

If the low potential cover is now touched with the finger, or otherwise connected with the earth, electrons flow from the low potential cover to the earth, thus leaving a deficit of electrons upon the cover, and raising its potential to zero. This forms the second case. If the connection with the earth is now broken and the cover lifted, it will be found to be positively electrified, inasmuch as electrons escaped from it when it was connected with the earth. Moreover, it has been lifted away from the low-potential region of the hard rubber plate, and is, therefore, now at high potential, thus forming the third case.

The only service performed by the negatively charged rubber plate is to lower the potential of the cover, so that electrons will escape from it when it is connected with the earth. Furthermore, all these operations have in no way disturbed the original negative electrification of the hard rubber plate; and so, after the cover has been discharged, it may again be lowered upon the plate and another charge secured in the same way. A single electrification of the hard rubber plate enables us, therefore, to obtain by induction an indefinite supply of electricity, simply by using this plate to lower the potential of the cover before this is connected with the earth.

The positive charges of electricity, which are thus secured each time upon the cover, may be used to charge a conductor, or may be consumed in performing mechanical work. Accordingly, each electrification of the cover represents a certain store of energy; and since the electrification of the hard rubber plate remains unchanged, the question may fairly be asked, what is the source of this energy? The answer lies in the fact that more work is required to lift the cover after it has been connected with the earth than before, by the amount of work required to overcome the attraction of the negatively charged rubber plate and the positively charged cover. In other words, the positive charge upon the cover has been lifted from a region of low potential to a region of higher potential, and, in so doing, every unit of this charge requires the expenditure of an amount of work which measures the difference of potential between the two positions, as we have seen under the measurement of potential (p. 356).

The positive charge, when lifted from the lower to the higher potential, possesses, in the latter position, an amount of potential energy equivalent to the work done against the electric forces in raising it to the higher potential. If left to itself, the electrified body upon which this positive charge rests would then move again towards the region of low potential, and thereby perform work; or, what is the same in effect, electrons would flow through a suitable conductor from a region of low potential to this electrified body at high potential, and perform the same amount of work. In either event, a current would be said to flow from the body at high potential towards the region of low potential. We now believe, however, that it is the electrons moving from the region of low potential towards the region of high potential that constitutes the current, though long-continued belief in the idea that it was the positive charge that moved compels us still to say that the current flows from the region of high potential towards the region of low potential, or in a direction opposite to that in which the electrons move.

327. Faraday's Ice-Pail Experiment. The great investigator, Faraday, performed a classic experiment which beautifully illustrates the flow of electrons under a difference of potential, and the redistribution of these electrons which one body can effect upon another by induction.

For lack of a better conductor, Faraday used a metal pail for carrying ice, and this he insulated by placing it upon a sheet of glass, at the same time connecting it by a fine wire with his gold-leaf electroscope. Into the ice-pail he then lowered a positively electrified ball held by a silk thread, and observed upon his gold-leaf electroscope the effect produced by the inductive action of the ball upon the ice-pail. His arrangement may be represented by the accompanying diagrams, Fig. 211. The ice-pail and the electroscope being originally in their neutral condition, the positively charged ball is lowered into the pail without touching its sides, and held by the silk thread. This positively charged ball then acts as the body *A*, and the ice-pail with the electroscope attached represents the body B of the previous sections.

If we think of the various phenomena as occurring because of the action of electrified particles upon one another, we can see that the positively charged ball when introduced into the ice-pail will effect a redistribution of the electrons upon the latter, attracting them to the inner surface of the pail, and leaving the outer surface thereby charged positively. The electroscope which is attached to the pail therefore indicates a positive charge. If the ball is moved to another position within the pail, the distribution of electrons upon the interior surface is altered so that they congregate in greatest numbers nearest the ball, but the exterior charge, as indicated by the electroscope, undergoes no change. If the ball now touches the inner surface, all traces of electrification upon the interior disappear, but the charge upon the exterior remains unaffected.

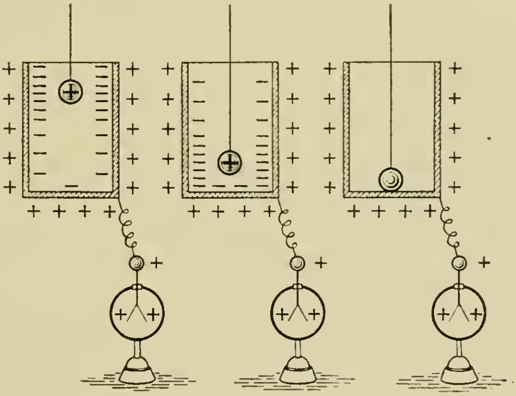

FIG. 211. The ice-pail experiment.

The explanation of these phenomena upon the potential theory is most helpful in enabling us to understand the flow of the electrons under a difference of potential. The ball, being charged positively and removed from all other electrified bodies, is at high potential, and this raises the potential of all other bodies in its neighborhood. The ice-pail therefore is raised in potential; and since the interior is nearer the region of high potential, it is for an instant at higher potential than the exterior. The free electrons upon the pail and the electroscope are then redistributed, flowing from the lower potential exterior to the higher potential interior, charging the interior negatively, and leaving the electroscope and the exterior of the pail charged positively. This flow of electrons continues until the electroscope and both surfaces of the pail are at the same potential. If the little charged ball is now moved about within the pail, but at no time touching its sides, not the slightest change in the deviation of the gold-leaves can be detected, showing that when the body *A* is entirely surrounded by the body *B*, the potential of *B* does not depend upon the position of *A* upon the interior.

The charge upon the interior of the ice-pail illustrates the remark made in a former paragraph (p. 346), that it is only when a charged body is actually brought inside another, but without touching it, that a static charge can exist upon the interior surface of a conductor.

If the charged ball is made to touch the inner surface of the ice-pail, the gold-leaves show absolutely no movement whatever; and if the ball, which then forms part of the inner surface of the pail, is carefully withdrawn, it will be found to be completely discharged. Moreover, the charge upon the interior of the pail has entirely disappeared, showing that the negative charge induced upon the interior is equal to the positive charge upon the little ball. The positive charge upon the exterior of the pail and upon the electroscope remains, however, this being exactly equal to the original charge upon the ball.

328. The Lines of Force in the Ice-Pail Experiment. The various electrical phenomena which the three stages of this experiment illustrate may be most beautifully studied by considering the lines of force which emanate from the charged ball when it is introduced into the interior of the ice-pail.

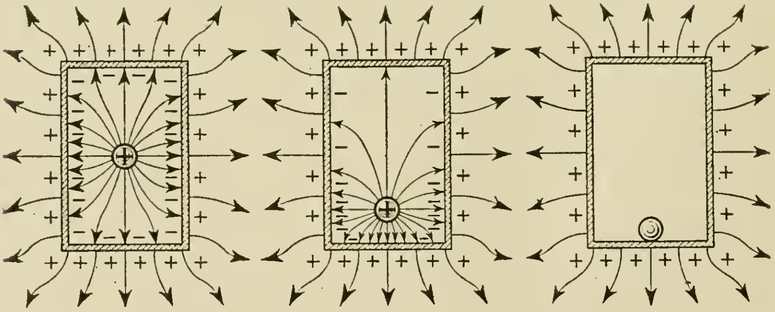

FIG. 212. The lines of force in the ice-pail experiment.

If the ball is lowered into the pail and the metal cover put on, all the lines of force which go out from the high potential ball are intercepted by the pail, and radiate again from the pail, showing that the exterior is positively charged. Moreover, the lines of force, as they go out from the ball and terminate upon the interior of the pail, show (p. 372), that the interior of the pail is charged negatively. If now, the ball is moved about in the pail, the lines of force, as they diverge from the ball, will congregate most thickly on the inside of the pail nearest the ball, but this will in no way affect the distribution of the lines of force which radiate from the outside of the pail.

When the little charged ball is touched to the inside of the pail, the high potential of the ball causes electrons to flow to it from the lower potential pail; and since all traces of electrification have now disappeared upon the interior of the pail as well as upon the ball, it is proved that the negative charge upon the interior of the pail was just equal to the positive charge upon the little ball. The ball now forms a part of the interior of the charged ice-pail, and, as we might have anticipated, it has lost its charge completely, since we know (p. 345) that, in general, there can be no electrification upon the interior of a charged body. It is to be observed, furthermore, that all traces of lines of force also have disappeared from the interior of the pail; and since lines of force always go out from a body of high potential and end upon a body of lower potential, there can be no difference of potential upon the interior of the pail. Therefore, the interior the pail is at a potential which is constant throughout, and which is the same in value as that of the pail itself (p. 365).

Omit from here

329. The Holtz Machine. The Holtz machine is a device for obtaining a continuous supply of electricity by induction, and might be called a continuous electrophorus. This consists of a fixed glass plate, denoted by the larger circle in the diagram, Fig. 213. and a smaller rotating glass plate nearer the observer, denoted by the smaller circle. Upon the back side of the fixed plate are pasted two paper or tin-foil sectors *A* and *B*, and upon the front side of the rotating plate are fastened a number of little tin-foil buttons. Two metal conductors *C* and *D* are bent around the edge of the plates and terminate on the front side of the revolving plate in metal brushes. As the plate turns, these metal brushes so touch the tin-foil buttons, that when the latter are just under the brushes they are connected with the sectors *A* and *B* upon the back side of the fixed plate. Two metal conductors *E* and *F* are supported near the front of the revolving plate, the ends near the center of the plate being rounded into knobs, and those near the circumference being provided with a series of sharp points, called the combs. These sharp points extend directly towards the moving plate and lie very near to it. This plate is supposed to turn towards the little metal brushes on the conducting wires *C* and *D*, and therefore in the direction indicated by the arrows.

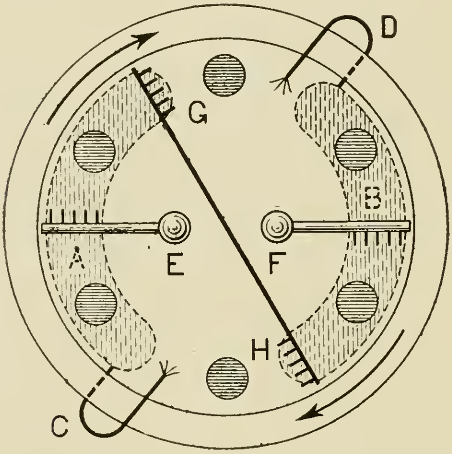

Fig. 213. The Holtz machine.

If we follow the conception that the electrical processes take place because of the mutual action of electrified particles, we may assume that a number of chance electrons have happened to accumulate upon one of the sectors, *A*, for instance, thereby charging it negatively. These electrons in *A* repel the electrons in the tin-foil buttons near this sector; and when these buttons pass under the points, or the comb, of the conductor *E*, electrons escape from the buttons and from the surface of the glass plate and flow through the points to this conductor, charging it negatively, and leaving the buttons and the plate charged positively. As the plate revolves, these positively charged buttons touch the metal brush at the end of the conductor *D*, thus enabling electrons to flow from the sector *B* through this conductor upon the buttons, making these once more normal, and leaving *B* positively charged.

Since *B* is now charged positively, it attracts the electrons in the conductor *F*, so that these flow through the points off upon the buttons and upon the surface of the plate, leaving *F* charged positively, and charging the buttons and the surface of the plate negatively. When a button, thus charged negatively, touches the tinsel brush on the conductor *C*, electrons from this button flow through this brush and the conductor to the sector *A*, charging this sector still more strongly negatively, and leaving the button once more in its normal condition. As this portion of the plate now turns so as to pass under the comb opposite the sector *A*, the electrons upon the buttons are once more repelled through these points to the conductor *E*, thus charging this conductor still more strongly negatively. The glass plate, moreover, still contains upon its surface the electrons that flowed off upon it when it was under the comb of the conductor *F*, near the sector *B*. These electrons upon the surface of the plate are likewise repelled by the electrons upon *A*, and therefore flow off from the plate through the points to the conductor *E*, thus increasing very greatly the negative charge which is already upon this conductor.

The effect of the entire revolution, which is now complete, has been to increase the negative charges upon *A* and *E*, and the positive charges upon *B* and *F*. Further revolutions serve to increase this effect until the charges in *E* and *F* become so great that the air is pierced, a spark passes, and electrons flow from *E* to *F*.

Calling to our aid the fuller explanation by the theory of potential, we see that, inasmuch as *A* is charged negatively and is uninfluenced by other electrified bodies, it is at low potential. Since *A* is at low potential, the potential of everything in its neighborhood is low, and therefore the potential of the conductor *E* is low. Furthermore, the end of *E* near *A* is nearer the region of low potential than is the farther end, so that, for an instant, this end is at lower potential than the remote end. The free electrons upon *E* now flow from the points towards the knob, leaving the points charged positively, and charging the knob negatively, though points and knob are at the same potential.

But a series of points upon or near a conductor always gives rise to a flow of electrons. Accordingly, when a button passes under these points, the low potential of *A* and the higher potential of *E* cause some of the electrons in the button to flow off upon the points, charging *E* negatively and leaving the button charged positively. Moreover, it is thought that the difference of potential between *A* and *E* disengages electrons from the surface of the glass, and these, in turn, flow likewise through the points to the conductor *D*, charging this conductor more strongly negatively, and leaving the surface of the rotating plate positively charged.

As the plate now turns, it moves away from the region of low potential produced by *A*, and so, together with the positive charge upon its surface, it becomes at high potential. Therefore, when a tin-foil button touches the metal brush at *D*, electrons flow from the sector *B* through the conductor *D* to the button, making the latter once more normal, and leaving the sector *B* positively charged, and so at high potential. The potential of the positive charge upon the plate is already high, and is raised still higher by being brought into the proximity of the high potential sector *B*. Accordingly, when the plate turns so as to pass the points on the conductor *F*, electrons flow from this conductor through the points upon the plate, thereby leaving *F* positively charged.

Moreover, it is under the influence of the high potential sector *B* that electrons flow from the conductor *F* off upon the plate, and therefore this flow is greater than it would be if the high potential of the plate were due only to the charge which it received when passing under the comb of the conductor *E*. As might be expected, then, when the plate passes out from under the influence of *B* and reaches its lowest position in the revolution, it is found to be charged negatively, and also to be at low potential. A low potential button now touches the brush on the conductor *C*, so that electrons flow from the button through *C* into the sector *A*, charging *A* still more strongly negatively, and reducing its potential still lower.

As the plate has now made one complete rotation, the above processes will be repeated with the next revolution. The action of the machine is therefore such that the sector *A* becomes lower and lower in potential, and the sector *B* higher and higher, the conductor *E* accumulates more and more electrons and sinks lower and lower in potential, while the conductor *F* loses more and more electrons and therefore rises higher and higher in potential, until the difference of potential between *E* and *F* is sufficiently great to cause the air to be punctured and a spark to pass. When *E* and *F* are thus discharged, the above processes must be repeated, until the conductors become again sufficiently charged to produce a spark between the knobs.

One further important feature of the machine must now be mentioned, namely, the diagonal conductor, *GH*. This rod is supported in front of the revolving plate, and is provided at its extremities with a series of sharp points which extend directly towards the plate, its sole object being to increase the length of the spark which takes place between *E* and *F*. To understand the action of this conductor, let us suppose the knobs *E* and *F* to be so far apart that no spark can pass between them under the existing potentials; the action of the machine, as above described, will then proceed until the conductor *E* has taken up all the electrons possible, and the conductor *F* has lost all the electrons possible due to the potentials which the machine can develop. Under these conditions *E* no longer takes up electrons from the plate as the latter passes the points, or comb, and so this part of the plate, with its electrons still upon it, passes by the points upon *E*, and comes to *G*. The remote end *H* of this conductor is at high potential, being in the neighborhood of the high potential sector *B*, hence the electrons upon the plate which pass by *E* flow off through the points to the conductor *G*, then flow downward through this diagonal rod to the high potential end *H*. Because of the points, however, these electrons cannot remain upon *H*, hence they flow through the points off upon the plate and the buttons, giving the lower portion of the plate an additional negative charge. The increased negative charge now upon the plate and the buttons compels still more electrons to flow from a button through *C* to the sector *A*, thereby causing the potential of *A* to fall still lower. This decrease in the potential of *A* lowers still further the potential of the negative charge upon the plate, so that now more electrons flow off upon E through the points, giving E a still greater negative charge, and lowering its potential still further. In an analogous manner the potential of F becomes still higher. Accordingly, the action of this diagonal comb is to increase the difference of potential between E and F, until finally this becomes so great that a spark passes, although the conductors are widely separated.

To [illegible]

CHAPTER XXV.

ELECTRICAL CAPACITY.

330. The Capacity of a Conductor. The previous sections illustrate the change which the potential of a body undergoes when other charged bodies act inductively upon it. When, however, a body stands alone and is uninfluenced by neighboring charges, the potential which the body then assumes depends upon the quantity of the charge with which it is electrified. If the body is charged positively, it is at high potential, and if it is charged negatively, it is at low potential. Moreover, if the quantity of the positive charge upon a given body is increased, the potential of the body will become higher in exactly the same proportion. Again, if equal charges are given to bodies of unequal size or form, the potentials to which these bodies are raised will, in general, be different.

Thus it is seen that the potential to which a conductor will be raised when a given charge is imparted to it, depends not only upon the magnitude of the charge, but also upon a property of the conductor, known as its *electrical capacity*. If the capacity is small, a given quantity will raise the conductor to a relatively high potential, but if the capacity is large, the same quantity will raise the conductor only to a relatively low potential. Since the potential to which a given quantity of electricity will raise a conductor is therefore directly proportional to the quantity, Q, with which the conductor is charged, and inversely proportional to the capacity, C, of that conductor, it follows that

$$V = \frac{Q}{C},$$

or

$$V = k\frac{Q}{C},$$

from which

$$C = k\frac{Q}{V},$$

where k is the factor of proportionality.

It still remains to define the quantity known as electrical capacity, and to decide upon an electrical capacity which we may adopt as unity. If, in the above equation, we agree to call the capacity unity when the quantity is unity and the potential is unity, then

$$k = 1,$$

and the equation becomes

$$C = \frac{Q}{V}.$$

Accordingly, the electrical capacity of a conductor is numerically equal to the ratio of the quantity of electricity with which the conductor is charged to the potential to which that quantity will raise the conductor. In other words, *the capacity of a conductor is defined as the quantity of electricity which will raise that conductor to unit potential.*

In the above expression we have, however, been able to give to k the value of unity only upon the single condition that we adopt a perfectly definite value for the unit capacity. Therefore, we have determined that *a conductor is of unit capacity when unit quantity of electricity will raise that conductor to unit potential.*

331. The Capacity of an Isolated Sphere. The capacity of an irregular conductor is, in general, difficult to determine, but the capacity of an isolated sphere admits of ready calculation.

We have seen (p. 365) that the potential V of a sphere is

$V = \frac{q}{R}$

where R is the radius, and we have learned above that the capacity C of a conductor is

$$C = \frac{Q}{V}.$$

Accordingly, the capacity C of a sphere becomes

$$C = \frac{Q}{\frac{Q}{R}} = R,$$

and therefore for a sphere

$$C = R;$$

that is, the capacity of a sphere is numerically equal to its radius. In other words, just as many units of electricity will charge a sphere to unit potential as the sphere has centimeters in its radius.

332. The Electrical Condenser. In a former section (p. 369) it was shown that the body A at high potential, Fig. 214, raises the potential of all other bodies in its neighborhood, so that when the neutral body B is brought into the presence of A the potential of B is raised. If B is now connected with the earth, electrons flow from the earth into B, and the potential of B falls to zero. But the lowering of the potential of B lowers the potential of all other bodies in its neighborhood, and accordingly lowers the potential of A. Therefore, if A had been charged to a potential as high as possible from a given source, its potential has now been lowered, and more electrons will flow from A to the same source. If A thus loses additional electrons, its potential is again raised to that of the source, and this raises the potential of B. Accordingly, if B is again connected with the earth, electrons will flow from the earth into B, thereby lowering its potential to zero, and likewise lowering the potential of A. If A is once more connected with the original source, a still further flow of electrons from *A* to the source, will now take place, thereby raising the potential of *A*, and at the same time raising the potential of the neighboring *B*, so that another flow of electrons can take place from the earth to *B*; and the above processes may be continued indefinitely.

If *B* is now connected continually with the earth, its potential is always zero and the potential of *A* will thus be kept low, so that great numbers of electrons will flow away from *A* to the source before the potential of *A* becomes equal to that of the source. By the operations here described, it is seen that a very large positive charge has been produced upon the body *A*, while electrons in correspondingly great numbers have been accumulated upon *B*, that is, *B* is charged negatively nearly as highly as *A* is charged positively. In other words, the simple presence of the body *B*, which is always maintained at zero potential, enormously increases the capacity of the body *A*.

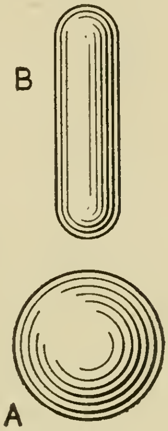

FIG. 214. The principle of the electrical condenser.

A system of two conductors, such as *A* and *B*, one of which is kept constantly connected with the earth, is known as an electrical condenser. The object of a condenser is to accumulate a very large quantity of electricity upon one of the conductors without raising its potential above that of the source. In other words, when the potential of the source is fixed, the condenser is a means of increasing the quantity of electricity upon one of the conductors at that given potential. But we have seen that each charge given to *A* is lowered in potential when *B* is connected with the earth. The condenser may also be regarded, therefore, as a means of lowering the potential of a given charge when the quantity is fixed.

333. The Leyden Jar. A common form in which the electrical condenser is found is the Leyden jar, Fig. 215. This consists of a glass jar coated inside and out for about two-thirds of its height with tin-foil. A conductor terminating in a knob projects above the mouth of the jar and makes possible a connection with the interior coating. Here we have the essentials of an electrical condenser, namely, two conductors separated by an insulating medium. The inner coating usually represents the body *A* in the above description of the condenser, while the outer coating represents the body *B*, because it may be more easily connected with the earth.

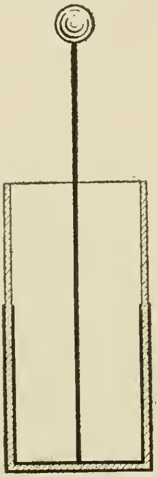

Fig. 215. The Leyden jar.

Although the jar is usually not insulated while being charged, we shall be able better to study the various operations of charging if we assume for the moment that it is insulated by being placed upon a sheet of glass or hard rubber. If we attempt to charge the insulated jar by connecting the knob with a high potential source, such as the Holtz machine or the electrophorus, we shall find, indeed, that a tiny spark will pass, indicating that electrons have flowed from the knob to the machine, but this flow will quickly cease, showing that a small positive charge is sufficient to raise the potential of the inner coating as high as that of the source. But the high potential of the inner coating raises the potential of the insulated outer coating; and so, if we now touch the outer coating, electrons will flow from the earth to this coating, thus lowering its potential to zero. This lowering the potential of the outer coating lowers also the potential of the inner coating, and another spark will now pass, showing that a second flow of electrons has taken place from the inner coating to the machine. This further loss of electrons from the inner coating once more raises the potential of this coating as high as that of the source, and the flow of electrons then ceases. The rise of potential of the inner coating raises for the second time the potential of the outer coating, so that again electrons will flow from the earth to this coating when the latter is touched, lowering its potential for the second time to zero. This lowering of the potential of the outer coating lowers once more also the potential of the inner coating, so that a third flow of electrons takes place from this coating to the source, or, in other words, it receives a third positive charge from the machine, and so on.

In practice, the jar is not insulated while being charged, but is connected by its outer coating with the earth, so that this coating is always at zero potential. Since this, in turn, keeps the potential of the inner coating low, large numbers of electrons can flow from this coating to the machine, or, what is the same in effect, this coating can receive a large positive charge. Indeed, the positive charge that accumulates upon the inner coating becomes at length so great, and the potential of this coating becomes finally so high, although it rises slowly, that a spark will pass, and electrons will flow from the outer coating to the inner, either over the top of the jar, or else through the glass, thus puncturing it and ruining the jar.

Whenever a spark passes between the knob of the jar and the source, that is, whenever electrons flow from the high potential inner coating to the machine, it is to be observed that an equal flow of electrons takes place from the earth to the outer coating. As this process is limited only by the strength of the glass or by the distance over the top of the jar, it is seen, either that the number of electrons which resided originally upon the inner coating must have been really very great, or else that the high potentials which are here involved are continually disengaging electrons from the atoms and molecules of the material of which the jar is made. Indeed, the electrons upon a body have never yet been entirely exhausted. After an indefinitely large number have been withdrawn from a conductor, it is always possible to obtain a still further flow by lowering the potential of the body below that of the surrounding region.

334. The Capacity of a Condenser. *The capacity of a condenser may be defined as the number of units of electricity that will raise one conductor of the condenser to unit potential while the other conductor is connected constantly with the earth.* It is far more difficult to determine the capacity of a condenser than of an isolated conductor, because it is found that the capacity of a condenser depends upon three things, namely:

1. Upon the size and to some extent upon the shape of the conductors.
2. Upon the nearness together of the conductors.
3. Upon the nature of the medium between the conductors.

The insulating medium between the conductors of a condenser is known as the *dielectric*. Thus in the case of the bodies A and B, referred to above (p. 389), the dielectric is the air, and in the case of the Leyden jar it is the glass. It is found that the capacity of a condenser depends very greatly upon the nature of the dielectric separating the two conductors. In the example just cited, where bodies A and B form the conductors of the condenser and air is the dielectric, it is found that A will raise the potential of B nearly three times as high if glass instead of air separates the two bodies. A piece of glass placed between these two bodies will produce the same effect as bringing B very much nearer to A, which indicates that induction takes place more readily through glass than through air. It is thought, indeed, that the glass offers an easier path for the lines of force than does the air.

This difference in the effectiveness of dielectrics in producing induction is due to a property of the medium known as the *specific inductive capacity*. *The specific inductive capacity of a dielectric is the ratio of the capacity of a condenser when this substance is used as the dielectric to the capacity of the same condenser if air were used as the dielectric.*

335. The Capacity of a Spherical Condenser. A spherical condenser consists of two concentric metallic spheres of radii r and r', separated by a layer of air or other dielectric medium, the outer sphere being connected with the earth, Fig. 216. If a positive charge of Q units of electricity is given to the inner sphere, the potential of the outer sphere will be momentarily raised, and electrons, equivalent in number to Q units of electricity, will flow to the outer sphere from the earth, thereby giving the inner surface of the outer sphere a charge of $-Q$ units. We are now to ascertain the potential to which the inner sphere is raised as a result of these processes.

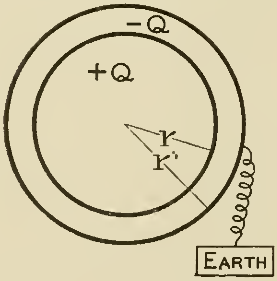

FIG. 216. A spherical condenser.

Since the potential of a charged sphere is constant throughout and equal to the quantity of electricity with which it is charged divided by the radius (p. 365), we see that the potential at all points within the outer sphere due to its charge alone is

$$\frac{-Q}{r'},$$

while the potential of the inner sphere due to its charge alone is

$$\frac{+Q}{r}.$$

Accordingly, the resultant potential V at all points within the inner sphere, and therefore the potential of the sphere itself, due to the superposition of these two potentials, is

$$V = \frac{Q}{r} - \frac{Q}{r'},$$

or

$$V = Q\left(\frac{1}{r} - \frac{1}{r'}\right).$$

But the capacity of a conductor (p. 387) is

$$C = \frac{Q}{V},$$

and therefore the capacity of this condenser is

$$C = \frac{Q}{Q\left(\frac{1}{r} - \frac{1}{r'}\right)},$$

$$= \frac{1}{\frac{1}{r} - \frac{1}{r'}}$$

$$= \frac{1}{\frac{r' - r}{r r'}},$$

$$= \frac{r r'}{r' - r}.$$

By adding and subtracting r^2 in the numerator, this expression may be given the form

$$C = \frac{r^2 + r r' - r^2}{r' - r},$$

$$= \frac{r (r' - r) + r^2}{r' - r},$$

$$= r + \frac{r^2}{r' - r}.$$

But (p. 388) r is the capacity of the inner sphere alone, regarded simply as an insulated conductor. The addition, therefore, of the outer sphere connected to earth increases the capacity by

$$\frac{r^2}{r' - r}$$

Moreover, this increase itself becomes greater the more nearly r' is equal to r, that is, the thinner is made the layer of air which separates the two conductors. It is seen, therefore, that the capacity of the condenser depends not only upon the size of the conductors, that is, upon the magnitudes of r and r', but also upon the nearness together of these conductors.

If some medium other than air separates the two spheres, this expression for the capacity of the condenser must be multiplied by the specific inducive capacity of that medium. The specific inducive capacity of air is taken as unity.

CHAPTER XXVI.

THE ELECTRIC DISCHARGE.

336. The Strain in the Dielectric. A phenomenon may be observed in the discharge of a Leyden jar which suggests the conclusion that when a condenser is charged the dielectric medium separating the conductors is under a severe strain. If a Leyden jar is discharged and allowed to stand for a few moments, it will be found that a second discharge may be obtained, feebler than the first, though still of considerable quantity. Another and another discharge may be obtained in a similar way, and in some instances as many as ten or twelve discharges may be drawn from the jar, each being much smaller than the preceding. This is thought to be due to a polarization and a possible turning of the molecules of the dielectric under the great difference of potential between the two coatings, after which the molecules, perhaps retarded by friction, return only slowly to their former positions when the potential gradient disappears after the first discharge. The charges carried by these polarized molecules constitute the second and the successive discharges, so that the phenomenon is probably not due to a "soaking" of the charge into the dielectric, as is sometimes said to be the case. Moreover, the theory of this strain in the dielectric receives support from the puncturing of the jar which frequently happens when the distance from coating to coating over the top is too great for a spark to pass. When the jar is thus broken, the glass has the appearance of being crushed, as if by great pressure.

Another interesting and curious phenomenon connected with the discharge of a Leyden jar confirms the theory that the dielectric medium plays a very important part in the discharge, and therefore probably also in the charge as well. If a jar with detachable coatings is charged and the coatings then removed from the glass, it will be found that these coatings may be touched, or connected together, or both connected with the earth, without producing even the slightest discharge. If, however, the coatings are replaced, the inner one always being lifted by an insulating handle, the jar will be found still to be charged as fully as before. If the walls of the glass jar are touched on opposite sides while the coatings are removed, little discharges from these particular areas will be obtained, proving that the surfaces of the glass are still highly charged. This apparently shows that the real electrification resides upon the surface of the dielectric instead of the metal coatings, and that the coatings serve simply as conductors, thus making possible the flow of the charge which resides upon the surface of the dielectric.

337. The Energy Expended in Charging a Conductor. It has been seen (p. 358) that the work done in carrying a unit of positive electricity between two points, A and B, at different potentials is

$$V_A - V_B ,$$

where V_A and V_B are the potentials at A and B respectively. When Q units of electricity are carried from B to A, the work done is

$$Q (V_A - V_B).$$

If B is at infinity, V_B is zero, therefore the work done in carrying Q units of electricity from a region where the potential is zero to a region where the potential is V is

$$QV.$$

Here it is assumed that the potential V does not change when the Q units are brought up. The energy, therefore, which Q units of electricity possess when brought up from a region at zero potential to a region where the potential is V is

$$QV.$$

The problem, however, is quite different if, as is usually the case in static electricity, we use these Q units, which are thus brought up, as a means of charging a conductor or condenser. Here the potential of the conductor is zero at the beginning, and rises as increments of electricity are added to it, but does not attain its final value V until the full number of Q units have been brought up. It is, therefore, only the last added increment of electricity that has been raised to the potential V, the others having been brought up to potentials intermediate between 0 and V. Accordingly, the *average* potential to which the entire number of the Q units have been raised is $\frac{1}{2}V$. From this it follows that the work which has been done in charging a conductor or condenser with Q units of electricity to the potential V is

$$Q \times \tfrac{1}{2}V,$$

or

$$\text{The energy of charge} = \tfrac{1}{2}\,QV.$$

Similarly, if a conductor is charged by Q units of electricity to the potential V, and is then discharged, only the first unit of electricity will flow out at this potential, since the potential of a conductor falls as successive increments of electricity are withdrawn. Other units leave at increasingly lower potentials, until the potential of the conductor is reduced to 0 as the last unit flows out. The *average* potential, therefore, at which the entire Q units have flowed out is $\frac{1}{2}\,V$, and the work which is thereby done by the discharge of a conductor or condenser is

$$Q\times\tfrac{1}{2}V.$$

Accordingly,

$$\text{The energy of discharge} = \tfrac{1}{2}\,QV.$$

338. The Lichtenberg Figures. If an unelectrified non-conductor, such as a piece of hard rubber, is touched by a positively electrified body, and then also touched at another place by a negatively electrified body, these charges will creep a short distance over the non-conductor, but no result can be detected until the plate is dusted over with a mixture of red lead and sulphur. As this mixture is sifted upon the plate, the friction between the sieve and the powder electrifies the red lead positively and the sulphur negatively; hence the positively electrified red lead is attracted to the negatively electrified spot upon the hard rubber plate, while the negatively electrified sulphur is attracted to the corresponding positively electrified spot. It is then found that the positive electrification of the plate has spread out in the most beautiful fern-like branching forms, while the negatively electrified spot seems to be charged in irregular and characteristic patches, the whole having a faint resemblance in appearance to a bunch of grapes.

By substituting an ordinary dry plate for the hard rubber sheet, Watts has succeeded in recording photographically the beautiful figures produced by electric charges, as these spread out and creep over the surface of the film. It is then seen, far more perfectly than by the Lichtenberg method, how the branching forms of the positive electrification extend out into the most delicate ramifications, while the patches produced by the negative charges are resolved into feather-like fronds of singular beauty and regularity. The reason why such different and peculiar effects are produced by the two electrifications is not known.

339. The Electric Spark. We have seen (p. 384) that the operation of an electric generator, such as the Holtz machine, consists in the transfer of electrons from one conductor to another, thereby raising the potential of the conductor from which they are taken and lowering the potential of the one into which they flow. The difference of potential thereby produced is so great that the intervening air or other dielectric is finally pierced, and electrons flow through the channel thus provided from the region of low potential to the higher, though we ordinarily speak of the current as flowing from the region of the higher potential to the lower. The bright spark which occurs, however, is not the electricity, but is a column of air heated to incandescence by the energy of the electric discharge. The crooked path of a long spark is due to the fact that the discharge always seeks the easiest path, and so deviates from a straight line to pass through particles of dust or traces of moisture which may be present.

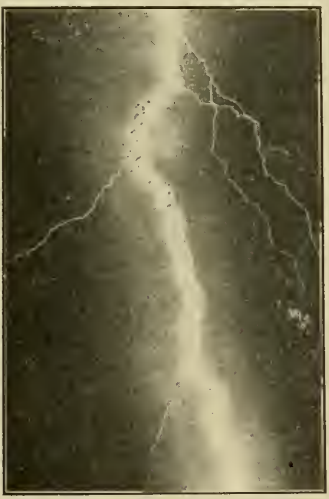

FIG. 217. Photograph of lightning by Kayser, showing banded appearance.

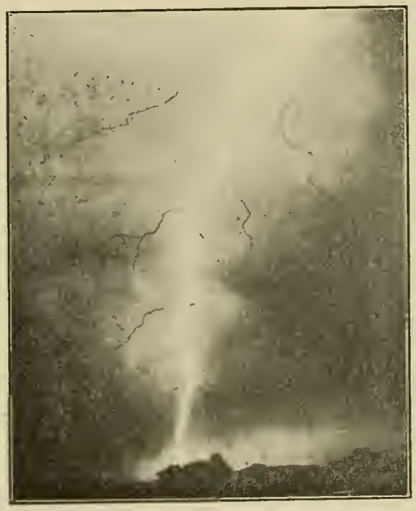

FIG. 218. Dark lightning.

The lightning flash is simply an electric discharge upon an enormous scale which takes place between two clouds, or between a cloud and the earth when these are oppositely charged. Certain phenomena connected with lightning, however, have as yet received no explanation. Prominent among these has been until recently the dark line or lines extending lengthwise through the flash, dividing the latter into a series of ribbon-like bands.

as seen in the accompanying flash, Fig. 217, photographed by Kayser. The phenomenon of "dark lightning," which is occasionally observed, Fig. 218, has been shown by Wood to be due to a photographic reversal.

If the Leyden jars are removed from the conductors of an electrical machine, the discharge takes place as a brush of bluish light which extends between the two conductors, and which has a form upon the positive end different from that upon the negative end. The brush seems to extend out from the positive end, while the negative end is surrounded by a curious dark space. A sharp point placed on or near one of the conductors will produce a silent glow-discharge of bluish light. Many of the phenomena of the discharge are not well understood.

340. Lightning and Lightning Rods. It is, indeed, upon the effect of points in producing a silent glow-discharge instead of a disruptive discharge that lightning rods serve in a measure as a protection from lightning. To be effective, a rod, which may well be simply a piece of stout telegraph wire, should be securely nailed to the building and be carried over the roof in several branches, finally terminating in a number of upright projections perhaps two feet high and not more than fifteen feet apart. The lower end of the rod, unless it can be connected to underground water pipes, should be riveted or soldered to a large metal plate, and should be deeply buried in moist earth, preferably with a load of coke above and below. Care should be taken to nail the rod firmly to the roof and to solder it rigidly to the metal eaves-trough and to the water spout. Insulators separating the rod from the building should never be used. The entire surface of the roof is thus maintained at zero potential, and a ready escape to the earth is provided for electricity if the rod should be struck. An even greater precaution would consist in covering the entire roof with a network of wire fencing and soldering this to the lightning rod, for then the building would be practically inside a metal conductor, within which no electrification can exist. A building provided with a metal roof effectively grounded could not be injured by lightning.

If a cloud charged either positively or negatively approaches the building, the difference of potential between the cloud and the earth will cause a silent glow-discharge to take place from the sharp points with which the rod is provided, and the difference of potential is thus, in general, so diminished that a flash does not take place. Even if a flash were to occur between the cloud and the building, it is most probable that the discharge would follow the stream of electrified air particles escaping from the sharp points and so be directed to the rod, when it would then pass harmlessly to the earth. Thus when a single cloud is above a building and a discharge takes place either in the silent glow or in the flash, it appears that the lightning rod is effective in protecting the building.

There is, however, so far as is now known, one condition under which a lightning rod is entirely without avail in affording protection, namely, if a flash takes place between two clouds when the lower one is directly over a building. In this event the lower cloud becomes instantly charged, and the potential rises with such rapidity that it cannot be equalized by the glow from the points, and so this cloud in turn immediately discharges itself disruptively to the earth. The flash then manifests all the characteristics of inertia and dashes to the earth with an impulsive rush, breaking through the air apparently at almost any point, totally neglecting to seek the path that offers the least resistance, and being quite as liable to strike the earth or a well-protected building, even at a place very near a lightning rod, as to follow the readier path which the rod offers. While a lightning rod, therefore, decreases the number of accidents, it by no means affords perfect immunity from injury caused by lightning.

As to the material of which the rod is made, it is now thought that a stout, galvanized iron wire is quite as good as a copper rod, and probably much better. In the following paragraph it will be shown that a flash of lightning, like every electrical discharge, is oscillatory in character, that is, that it consists of many surges back and forth, until the energy of the flash is finally converted into heat. In an iron wire, which may have a resistance even greater than a rod of copper, there is a larger mass of metal present, and therefore less liability for an undue heating or possible melting of the rod to take place. Indeed, the resistance of the rod is of great advantage, inasmuch as it dissipates into heat that energy of the discharge, so that the flow through the rod may perhaps take place as a single surge, instead of many oscillations back and forth. Furthermore, as an oscillatory discharge is confined almost entirely to the surface of the conductor, a lightning rod should have as large a surface as possible, and would be more effective if it were rolled out in the form of a flat strip, or sheet, than if the same amount of metal were kept in the form of a solid rod.

341. The Oscillatory Character of the Spark. In the previous sections it has been shown that the discharge of an electrified conductor or a charged condenser represents a flow of energy. This energy of the discharge is made manifest by the heating, and sometimes vaporizing, the conductor through which the flow takes place, and in the amount of mechanical work which this discharge is able to do, such as piercing poor conductors and even glass. The energy which is stored up as potential energy in the charged body becomes kinetic in the discharge, and is found to manifest characteristics not unlike those associated with kinetic energy in matter.

Possibly the most obvious property of matter, certainly as regards its motion, is inertia (p. 99), and it would seem, at first, that surely the energy of the electric discharge could not possess a characteristic which would give to it a property in any way analogous to inertia, but, such, indeed, is found to be the case. It was predicted mathematically, and afterwards verified by experiment, that just as the energy represented by a swinging pendulum varies from potential to kinetic and from kinetic to potential, until it is finally dissipated by friction, resistance of the air, etc., so the potential energy represented by the charged conductor will become kinetic as the discharge passes, and this rush of electrons from one conductor to another will cause an accumulation of electrons upon that other conductor. This same process then again takes place but in the reverse direction, and so on, until the entire energy of the discharge is consumed in heating the air, producing noise, doing work, etc.

The discharge of a conductor, therefore, very closely resembles the vibration of a pendulum, and the electrons pass first one way and then the other, back and forth many times along the column of heated air which we call the spark. In other words, as soon as the electrons constituting the charge are set in motion, they seem to possess a characteristic in every way analogous to inertia, so that what we see as a single spark is really a hundred or more oscillations or surgings of this store of energy, which dashes like a great wave across the gap represented by the spark, and collects upon the farther conductor only to surge back again in the opposite direction, and to repeat this operation many times.

Since this entire effect is accomplished in a fraction of a second, it cannot under ordinary conditions be detected by the unaided eye. If, however, by means of a rapidly rotating mirror we throw an image of the spark upon a screen, the spark, which would be drawn out into a line if it were continuous and of finite duration, is seen to consist of a great number of separate flashes taking place first in one direction and then in the other, showing that it is discontinuous and oscillatory in its character. Feddersen of Leipzig was the first to show these oscillations experimentally. By allowing the image of the spark to fall upon a rapidly-moving photographic plate, he found that each oscillation had there recorded itself, so that what seems to the eye to be a single spark is really a great number of oscillations of the discharge, taking place in opposite directions.

An undoubted explanation of the banded, or ribbon-like appearance of the lightning flash (p. 399), which has always been supposed to be connected with the oscillatory character of the discharge, has recently been given by Larsen. By mounting a photographic camera in such a way that it could be rotated about a vertical axis, this young Dane has succeeded in photographing a flash of lightning as remarkable as the one shown in Fig. 219, which is here reproduced by courtesy of the *Scientific American*. Had the camera been stationary, the flash, in all probability would have been recorded as a single bright line, or at most as two or three bright lines, separated by corresponding dark intervals. Here, however, since the photograph is so drawn out laterally that at least forty distinct discharges may be recognized, it is evident that the flash consisted of a series of oscillations, and that in the interval between discharges the camera had turned through an angle sufficiently great to bring the image of each oscillation upon a different portion of the plate. From the measured width of the photograph and from the known velocity of the camera, the entire duration of the flash was determined to be a little less than half a second. Since it is thus found that a flash may persist for so long an interval of time, the successive oscillations probably account also for the peculiar flickering appearance occasionally observed in a flash of lightning.

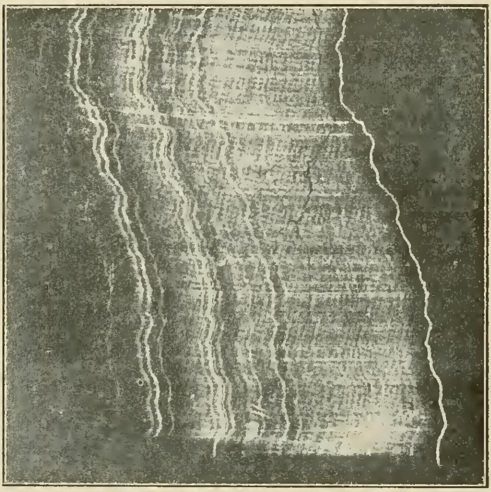

Fig. 219. Lightning photographed with revolving camera showing oscillatory discharge and dark flash.

The oscillatory character of the discharge furnishes likewise an explanation of the series of dark lines paralleling the bright ones which have been photographed with a stationary camera, as in the famous picture by Kayser, Fig. 217, since it is now known that the flash may have duration of nearly half a second. If a wind were blowing at the time, it is probable that the irregular column of heated air, through which the flash pierced its way at the first discharge, may have drifted to an appreciable extent before the second oscillation occurred. The returning flash, which took place through the same column of heated air, would then follow a path in every way parallel to its former course, but at some little distance from it, and so would produce an image in a slightly different position upon the photographic plate. The third and the successive oscillations would produce images likewise displaced, and the banded appearance of the flash would thus be produced.

342. The Propagation of Electrical Vibrations in the Ether. It has been shown (p. 364) that the ether all about a charged body is thrown into a state of strain. It is seen, therefore, that when a spark of electricity passes, the ether in the neighborhood is strained first in one way then in the other, due to the oscillations. Accordingly, the spark is a center from which spherical waves in the ether go out, very much as circular waves upon a surface of water proceed from a disturbance which takes place at the center. It is readily seen, then, that when a Leyden jar, for example, is discharged, the ether all about that jar is vibrating in unison with the oscillations which take place back and forth in the spark. These oscillations, and also therefore, the character of the vibrations in the ether, depend upon the capacity of the jar and upon the length and condition of the conductor (self-induction) which connects the two coatings.

If, therefore, another jar similar to the first is placed near, and the capacity of this second jar is carefully adjusted to be the same as that of the first, these ether vibrations which are taking place around this second jar are then exactly the same in nature and period as those which a spark from this jar would send out if it itself were discharged. Accordingly, if a wire connects the outer and inner coatings of this second jar, similar in every way to the circuit connecting the outer and inner coatings of the first jar, except that the circuit of the first jar is provided with a spark gap, while the second one is not, surges will take place in the second jar similar to those in the first. If, then, a conductor is brought from the inside coating of the second jar up over the edge of the jar and down on the outside almost to the outside coating, a little spark will pass across this short gap on the second jar every time a spark passes between the coatings of the first jar.

This is a beautiful example of sympathetic vibrations, whereby a vibrating body tends to generate vibrations in all other bodies which have a period of vibration equal to its own. Indeed, this experiment with the two Leyden jars is spoken of as illustrating "electrical resonance." It is due to these waves in the ether, which go out in all directions from an electric spark, that it is possible to telegraph between distant places without wires.

MAGNETISM

CHAPTER XXVII.

MAGNETIC PHENOMENA AND THEORIES.

343. Fundamental Phenomena. A magnet may be defined as anything that will attract small particles of iron. In this definition are included two classes of magnets, namely, natural magnets and artificial magnets. A natural magnet, or lodestone, is a variety of iron ore which is found to be magnetic and which retains its magnetism after it is taken from the ground. An artificial magnet is a bar of iron or steel which was originally non-magnetic, but to which this magnetic property has been artificially imparted. In the year 1600 Gilbert found that artificial magnet could be made simply by rubbing bars of iron or steel with a piece of lodestone, after which these bars would manifest all the properties of the natural magnet. He also found that a certain difference exists between the magnetization of a bar of soft iron and a bar of steel. A bar of soft iron remains a magnet and possesses all magnetic properties as long as it is in contact with the magnetizing substance, but loses more or less completely all traces of magnetism when the lodestone is removed. A bar of hardened steel, on the other hand, retains its magnetism permanently, and is, therefore, said to be more *retentive* than soft iron, or to possess greater *retentivity*.

If a natural magnet is dipped into a mass of iron filings, the filings will be seen to be attracted to the lodestone in great profusion. It will be noticed, however, that they adhere in greatest numbers to the edges and corners, and that they seem to be concentrated chiefly on two opposite sides of the piece of lodestone.

while the smoother and more central portions are practically devoid of filings. Moreover, it will be observed that the filings are not simply attracted in confused masses, but that they are all attached end to end and form little tufts which stand out from the angular portions of the magnet in a most beautiful way. Similarly, if a permanently magnetized bar of steel is plunged into a mass of iron filings, the filings will form the same remarkable tufted arrangement about the ends of the bar, but will diminish in quantity farther away, and will not be attracted at all at the center. The magnetism of the bar, therefore, seems to be concentrated at or near the ends of the magnet at points which are called the *poles*.

The poles of a magnet, accordingly, are the points near the ends from which the attractive force seems to emanate. The straight line joining the poles is called the *axis* of the magnet, and the distance between the poles is known as its *length*, the length of the magnet being, in general, about one-twelfth less than the length of the bar of steel. Inasmuch, however, as the location of the poles is somewhat uncertain, the length of the magnet, that is, the distance between its poles, cannot be exactly determined.

If a magnet is suspended by a thread or supported upon a pivot so as to turn freely in a horizontal plane, it will be found to assume a direction which is almost exactly north and south, one particular end always pointing north, while the opposite end always points south. If the magnet is displaced from this north and south direction, it will, after a few vibrations, settle back into its former position, the same ends again pointing north and south as before. The pole of the magnet, therefore, which points to the north is called the north-pointing pole, or more simply, the north pole, and that which points to the south is called the south-pointing pole, or the south pole. If the north pole of one magnet is presented to the north pole of another freely suspended magnet, the two will be found to repel each other. In the same way, if two south poles are brought near each other, they will be found likewise to repel; but if the north pole of one magnet is presented to the south pole of another, the two will be found to attract. We have, therefore, the first great law of magnetic attractions and repulsions, namely: *Like magnetic poles repel each other, while unlike magnetic poles attract.*

344. The Force Between Two Magnetic Poles. It is seen that the law for the mutual action of two poles is analogous to the law according for the mutual attractions and repulsions of electrified bodies. Since, however, it is impossible to speak of *quantities* of magnetism as we speak of quantities of electricity, the expression *strength of pole*, or *intensity of pole*, is used to express the ability which one magnetic pole possesses of exerting a force upon another magnetic pole when the two are brought near together. When one pole acts upon another to attract or repel it, a force is developed in the medium between the two. Coulomb found by experiment that this force is proportional to the product of the strengths of the poles and inversely proportional to the square of the distance between them. Thus, if F represents the force in dynes with which one magnetic pole of strength m attracts or repels another magnetic pole of strength m', where r is the distance between them, then

$$F : \frac{m\,m'}{r^2},$$

or

$$F = k\,\frac{m\,m'}{r^2}\;,$$

where k is the factor of proportionality.

This relation, which Coulomb verified by the use of the torsion balance, is known as Coulomb's law, and forms the second of the two great laws according to which mutual attractions and repulsions between magnetic poles take place.

345. The Unit Magnetic Pole. As yet, however, we have not determined the magnitude of the pole of unit intensity: but since force and distance (F and r) in the above equality may be expressed in terms of absolute units, so also may m and m' be likewise so expressed, provided only that k be eliminated. In order that k may have the value of unity, and so disappear from the equation, we agree to call m unity when m' is also unity, that is, when $m = m'$, and when F and r are each equal to unity, in which case

$$k = 1,$$

and the above equation becomes

$$F = \frac{m\,m'}{r^2}.$$

In thus assigning to k the value unity, we have, however, determined the magnitude of the pole of unit strength. Accordingly, *a pole of unit strength or intensity is a pole of such strength that it will repel a similar and equal pole at the distance of 1 centimeter with a force of 1 dyne.* From this it follows that if a pole of unknown strength were to exert a force, for instances, of 1000 dynes upon a unit pole when the two are 1 centimeter apart, the unknown pole would have a strength of 1000 units.

346. The Molecular Theory of Magnetism. If we break a long steel magnet into two parts, it will be found that it is not possible to have one half of the original magnet containing the north pole only and the other half the south pole only, but that each half will be a complete magnet possessing a north and a south pole of its own, Fig. 220, a new pole having developed at the fracture on each piece. If each of these pieces is again broken, the same will be found to hold true; and if the whole magnet is broken into little pieces, each fragment will be found to be a complete magnet possessing a north and a south pole.

If we imagine this process to be carried on indefinitely, we should ultimately reach the molecules, and we should then undoubtedly find that each molecule of a bar of steel is a complete magnet in itself, possessing a north and a south pole. In the magnetized steel the north poles of the molecules are thought, in general, to point one way and the south poles to point the other, so that we have what is equivalent to a great north pole at one end and a great south pole at the other. In the center of the bar the same orderly arrangement is believed to exist, but the effect of a row of south poles across the bar, as at *ab*, Fig. 221, is neutralized by the presence of a row of north poles close at hand, and thus no external result is produced. When the magnet is broken, however, these rows of opposite poles are separated from each other, and hence produce resultant poles on each of the fragments. A new pole, therefore, has not been created by breaking the magnet, but simply the disturbing and neutralizing effect of an equal and opposite pole infinitely near has been removed.

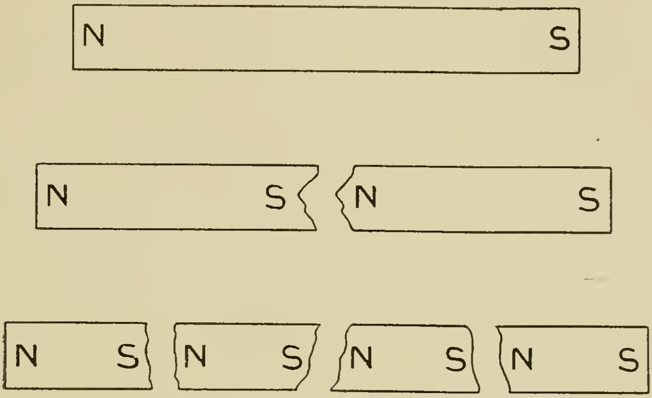

FIG. 220. Each fragment of a magnet is a complete magnet.

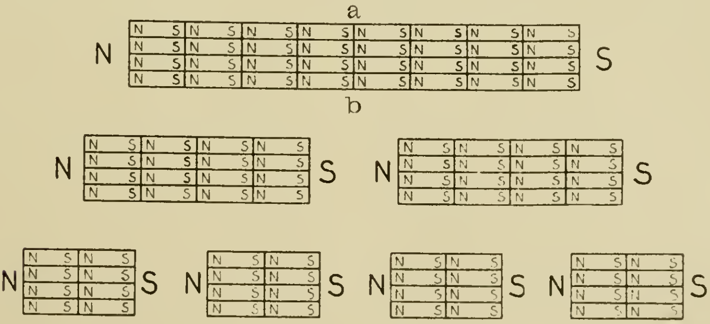

FIG. 221. The molecules of a magnet are magnetized.

When a bar of iron or steel is magnetized by rubbing with a permanent magnet, it is found that this magnet loses none of its own strength while imparting this remarkable property to the unmagnetized bar. This fact, combined with the inevitable conclusion that the molecules of a magnet are themselves magnetized, leads us to believe that the molecules of an unmagnetized bar of soft iron or steel are all magnetized, each having a north and a south pole, but that, owing to the mutual attractions and repulsions of these molecular poles, no definite order of arrangement exists. The molecules of an unmagnetized bar thus form a confused mass, Fig. 222, the poles pointing in every conceivable direction. Each molecule attaches itself to its neighbor in such a way that the north pole of one molecule attracts the south pole of an adjacent molecule, so that the molecules of an unmagnetized bar arrange themselves in irregular groups, each aggregation, however, representing the greatest molecular stability. The process of magnetization consists, therefore, not in magnetizing the molecules, but simply in giving order to this confused arrangement, that is, in turning the molecules around, so that their north poles will all point in one direction, and their south poles in the opposite direction.

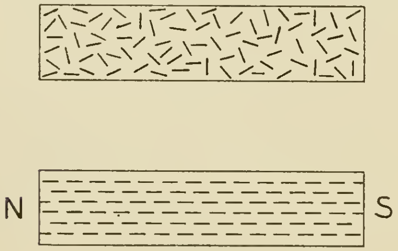

Fig. 222. The molecules in an unmagnetized bar of iron or steel and in the same bar magnetized.

In thus giving a uniform direction to the molecules, the marked distinction between soft iron and steel becomes at once apparent. It is found that the molecular forces which bind together the molecules of iron are not so strong as are those which exist between the molecules of steel, and, therefore, it is less difficult to magnetize a bar of soft iron than a bar of steel. In other words, for a given magnetizing force a bar of soft iron will become more strongly magnetic than a bar of steel; but owing to the greater freedom of motion among its particles, the soft iron loses its magnetic properties as soon as this force is withdrawn, while steel retains its magnetism. This simply means that the molecules of soft iron turn about and resume a thoroughly confused arrangement as soon as the magnetizing force is removed, while the molecules of steel, because they are held more firmly together, remain permanently in the positions which they assume when magnetized. It is more difficult, therefore, to magnetize a bar of steel than a bar of soft iron, but the steel retains its magnetism longer than the iron. This property of molecular rigidty which steel possess in greater measure than soft iron is known as *retentivity*. Steel, therefore, has greater retentivity than soft iron.

347. Phenomena due to the Molecular Magnets. There are many phenomena connected with magnets which confirm most conclusively the molecular theory of magnetism here presented. If a bar of steel is gently tapped while being magnetized, the jar thus produced gives the molecules greater freedom of motion and enables them more completely to set themselves parallel to one another, thus producing a stronger magnet. Immediately after the magnetizing force has been removed, the magnet loses somewhat in strength, since a number of the little molecules turn around notwithstanding the great retentivity of the steel. Hammering or pounding with a permanent magnet will weaken the magnet, because this aids the little molecules to turn around in obedience to their own attractions. Heating a magnet red-hot will cause it to lose all traces of magnetism, owing to the greater freedom of motion which the particles have at the higher temperature.

Whenever a strong magnet is quickly made, as, for instance, when a current of electricity is passed through a coil of wire containing a bar of soft iron, a faint click, called the "magnetic tick," is heard, which is thought to be due to something analogous to friction between the molecules as they turn and assume the more orderly molecular arrangement of a magnet. If an alternating current is sent through the coil, the molecules of the iron will set themselves first in one direction and then in the other, and the "tick" made by the molecules in turning occurs as frequently as the current reverses. This gives rise to a musical note, the pitch of which corresponds to the number of alternations of the current. Moreover, under these circumstances the iron grows warm, showing that energy is expended in overcoming the molecular friction as the molecules point first one way and then the other; and as might be expected, steel becomes heated in this way more than iron. This explains the "singing" of the transformer.

348. Limit of Magnetization—Magnetic Saturation. Furthermore, if the molecular theory is correct, we should anticipate that there must be a limit to the magnetization which a bar of iron or steel can acquire. Since in the unmagnetized bar the molecules are in confusion and lie in all possible directions, the process of magnetization consists so in turning these molecules that as far as possible they may all point one way.

We believe, then, that the strength of pole of a magnet, other things being equal, depends upon the relative number of these molecules which thus assume an orderly direction. The more nearly parallel to one another the molecules lie, the greater becomes the strength of the magnetic pole so produced. Therefore, as the magnetizing force increases, the molecules are made to lie more and more nearly parallel to one another, and the strength of the pole likewise increases. A limit, however, is finally reached when practically all the molecules which are free to turn lie parallel to one another, after which even a great increase in the magnetizing force produces at best only a feeble addition to the strength of the pole, and the magnet is said to be *saturated*. In a saturated magnet the limit of arranging the molecular magnets parallel to one another has been reached, so that the strength of a magnet cannot be increased beyond its point of saturation.

349. Residual Magnetism. When the magnetizing force is removed from a bar of soft iron, nearly all the molecules form groups again in obedience to their mutual attractions, owing to the small retentivity of the iron. A few molecules, however, remain pointing in the same direction, so that a feeble degree of magnetization remains, which gives rise to what is known as residual magnetism. Residual magnetism is the magnetism remaining in the bar of iron after the magnetizing force has been removed. The magnetism of a permanent steel magnet is thus entirely residual, the great retentivity of the steel making possible a large amount of residual magnetism.

350. Magnetic Moment of a Magnet. It is found that the ability of a suspended magnet to turn under magnetic influence is due not only to the strength m of its poles, but to the distance l between them. The product of the strength of one of the poles of a magnet and the distance between them is called the magnetic moment of the magnet. Accordingly, if M denotes the magnetic moment,

$$M = ml.$$

Since the location of the poles within a magnet cannot be precisely ascertained, the distance l cannot be exactly measured. It is, therefore, far easier to compute the magnetic moment of a magnet, which indeed, involves both m and l, than to determine separately either its strength of pole m or its length l.

351. The Intensity of Magnetization. The intensity of magnetization of a magnet is defined as the ratio of the magnetic moment to the volume of the magnet, or, in other words, it is the magnetic moment per unit of volume, and is usually denoted by I. That is,

$$I = \frac{\text{Magnetic moment}}{\text{Volume of magnet}}.$$

If a denotes the area of the cross-section of the magnet, that is, the area of the end of the magnet, and l the length of the magnet, then

The volume of the magnet $= al$;

and since

The magnetic moment of the magnet $= ml$, it follows that

$$I = \frac{m}{a}\frac{l}{l'}.$$

or, approximately,

$$I = \frac{m}{a},$$

since the length of the magnet may be regarded, roughly, as the length of the bar of steel. The intensity of magnetization, therefore, may also be defined as the strength of pole per unit of cross-section.

The interpretation of this expression is as follows: The effect of a magnet proceeds from its pole and depends upon its strength of pole. This pole, however, is not situated on the surface, but is a short distance within the magnet, and the magnetic influence which emanates from the pole must pass through the end of the magnet, and so be distributed over a surface equal to the area of the end. The effect, therefore, which proceeds from 1 square centimeter of the end, the area of which is a, is $\frac{1}{a}$ of that which proceeds from the pole m.

Accordingly, if two magnets have equal lengths and equal strengths of pole but different cross-sections, Fig. 223, the same total effect must come through the end of each magnet. The effect *per square centimeter,* however, is greater in the smaller magnet, indicating that in this the molecules have been rendered more nearly parallel, that is, that the iron or steel has been more highly magnetized, than in the larger. In other words, the two expressions, *intensity of magnetization,* and *strength of pole per unit of surface,* are identical in meaning,

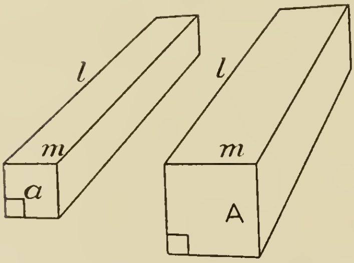

FIG. 223. The strength of pole per unit of surface is greater the smaller the cross-section.

since both refer to the degree to which the molecules of the magnet have been orderly arranged.

Since the magnetic effect due to the pole of strength m is thus distributed over the surface a, which is the area of the end, the expression

$$\frac{m}{a}$$

has also received the name magnetic density. This term, which is an unfortunate one, undoubtedly had its origin in the old idea that magnetism was a fluid, and that the quantity of magnetism m was spread over the surface a of the end of the magnet. Accordingly, this expression represents the amount of magnetism per unit of polar surface, that is, the magnetic density.

We see, therefore, that the intensity of magnetization is a measure of the magnetic state of the iron or steel, and must be thought of as a property of the entire body of the magnet or any part of it. If a magnet is cut into short lengths, Fig. 224, new poles, each equal in strength to the old ones, will be developed at the cut surfaces. Since the lengths are thus shorter, the magnetic moment of each piece is smaller, but so, also, is the volume, in the same proportion. If the magnet is split lengthwise into smaller sections, Fig. 225, the strength of pole of each piece will be less in proportion to the diminished area of the cross-section. Here also, since the strength of pole is less, the magnetic moment of each section is smaller, but only in exactly the same proportion as the volume has become smaller.

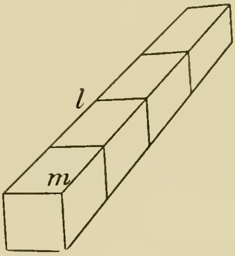

FIG. 224. The magnetic moment is proportional to the volume.

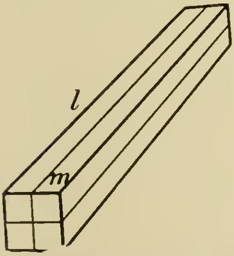

FIG. 225. Here also the magnetic moment is proportional to the volume.

In either event, the magnetic moment, $m\,l$, has become smaller, but so has the volume, $a\,l$, likewise become smaller and in the same proportion, and therefore the intensity of magnetism.

$$\frac{ml}{al},$$

or, approximately,

$$\frac{m}{a},$$

has remained constant. Accordingly, the ratio of the magnetic moment to the volume, that is, the magnetic moment per unit of volume, or the intensity of magnetization, *is a property of the magnet,* and is a measure of the degree to which the molecules of the iron or steel have been arranged so as to lie parallel and in the same direction.

While the three expressions, *intensity of magnetization, strength of pole per unit of surface*, and *magnetic density*, all refer to the same physical conception, the first of these, namely, *intensity of magnetization*, carries with it an idea of the inner magnetized state of the magnet which the others fail to convey.

CHAPTER XXVIII.

THE MAGNETIC FIELD.

352. Magnetic Potential. We have already found (p. 409) the expression for the force with which one magnetic pole acts upon another pole when the two are a certain distance apart. In the solution of many problems, however, it is simpler to use the expression for the magnetic potential at a point instead of the force at that point, meaning by the magnetic potential at a point the potential energy which a unit north magnetic pole would possess if placed at that point. Accordingly, the magnetic potential at a point is defined as the number of ergs of work required to bring the unit north magnetic pole from infinity to that point; and the difference of magnetic potential between two points is the work done in carrying the unit north magnetic pole from one point to the other. Two points have a difference of magnetic potential of one unit between them when one erg of work is required to carry the unit north magnetic pole from one point to the other. By a process of reasoning similar in every way to that employed in determining the electric potential at a point (p. 357), the magnetic potential V at a point at a distance r from a pole of strength m is

$$V = \frac{m}{r},$$

and the difference of potential between two points A and B at the distances r and r' from the pole m is

$$V_A - V_B = m\left(\frac{1}{r} - \frac{1}{r'}\right)$$

Moreover, the potential at a point due to the presence of several poles is the sum of the potentials which each pole would produce separately at that point. Thus, if a point A is distant $r, r', r'', r''', \ldots\ldots\ldots\ldots$ from poles $m, m' m'', m''', \ldots\ldots\ldots\ldots$ the potential at A is

$$V_A = \frac{m}{r} + \frac{m'}{r'} + \frac{m''}{r''} + \frac{m'''}{r'''} + \cdots\cdots\cdots$$

It must not be forgotten that if positive work is done in bringing up a unit north magnetic pole into the presence of the north pole of a magnet, we must think of the potential as being high in the neighborhood of a north pole. On the other hand, if a unit north magnetic pole were brought up near the south pole of a magnet, the two would tend of themselves to move together, hence a negative amount of work would have to be done, and therefore the potential due to a south pole must be regarded as low.

353. Equipotential Surfaces. We have defined the magnetic potential at a point to be the number of ergs of work required to bring a unit north pole from infinity to that point. We have also seen that if a unit north pole is brought up to a distance r from a pole whose strength is m, the potential is

$$\frac{m}{r}.$$

The same number of ergs of work would have to be done had the unit north pole been brought up from infinity along any other path to the same distance from the pole, and therefore through the points at the distance r from the pole a surface may be laid such that the potential at every point of this surface is constant. Such a surface is called an equipotential surface.

An equipotential surface, therefore, is an imaginary surface drawn around a magnetic pole or a number of magnetic poles, over the entire extent of which the magnetic potential has a constant value. At another distance from the pole another surface may be laid, at every point of which the potential is likewise constant, though of different value from that upon the first. In this way the entire region about a magnetic pole or a number of poles may be divided into a series of equipotential surfaces. These equipotential surfaces have a difference of potential of one unit between them when one erg of work is required to carry a unit north magnetic pole from one surface to the other. The forms of the equipotential surfaces as they exist about a magnet having two poles are far from simple.

However complicated an equipotential surface may be, its potential is everywhere constant, and therefore no work is required to carry a unit north pole between two points upon an equipotential surface. Moreover, the work done in carrying a unit north pole from one equipotential surface to another is independent of the path, since this path may always be resolved into two components, one of which is parallel to the equipotential surfaces, and the other perpendicular to these surfaces. Since no work is performed in carrying a magnetic pole parallel to an equipotential surface, the entire work is done in carrying this pole along a line perpendicular to these surfaces. In other words, the magnetic force always acts at right angles to the equipotential surfaces, or, equipotential surfaces and lines of force (p. 422) are always perpendicular to each other.

354. The Magnetic Field. The region all about a magnet within which the magnetic effect may be detected is called the magnetic field, or the field of force. The existence of this field of force about a magnet is thought to be due to a distortion of the ether in the neighborhood of the magnet in a manner analogous, though not similar to the distortion produced by an electrified body (p. 364). That is, the simple presence of a magnet is thought to disturb or strain the ether all about a magnet; and the influence which one magnet can exert upon another is due to this distorted state of the ether, or, in other words, to the presence of the two magnetic fields. There are certain phenomena which would seem to indicate that this distortion of the ether about a magnet is of the nature of a twist, or whirl, but this is not as yet proved. When a magnetic pole is brought into a field of force it is influenced by this distorted state of the ether, and so is acted upon by a force. While the distortion or strain of the ether exists in a field of force before a magnetic pole is brought into the field, *a force does not exist until a magnetic pole is actually there. A field of force, therefore, is not a force, but is that condition of the region about a magnet in virtue of which a magnetic pole is acted upon by a force when it is brought into the field.*

355. The Characteristics of the Field—The Direction of the Field. It is necessary to specify two characteristics of a magnetic field of force in order to define the field. These characteristics are the *direction* of the field and the *strength*, or the *intensity*, of the field.

The direction of the field at any point is defined as the direction in which a north magnetic pole would move if placed at that point. Thus, if *o*, Fig. 226, represents the north pole of a small imaginary magnet in the field of force of the magnet *NS*, *o* is repelled by the north pole of the magnet along the line *oa*, and attracted by the south pole of the magnet along the line *ob*. The resultant of these two forces is *oc*, which is the direction in which the pole *o* would move. The direction *oc* indicates, therefore, the direction of the resultant force at the point *o*.

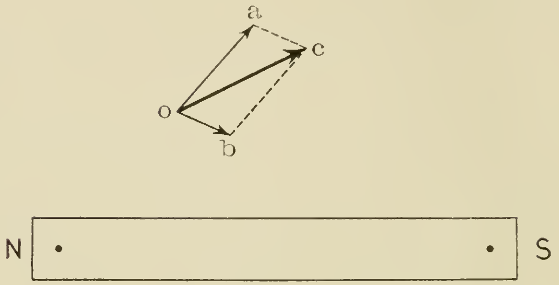

Fig. 226. The direction of the resultant force in a magnetic field.

356. The Lines of Force. Our study of the properties of a field of force is assisted greatly by the conception of *lines of force*. *A line of force in a magnetic field is a continuous line so drawn that it shows at each point the direction of the force at that point.* In other words, *a line of force in a magnetic field is a line along which a north magnetic pole would move when brought into the field.* Accordingly, if a small north pole, Fig. 227, is brought near the north pole of the magnet and left to itself, it will move, being repelled at each point of its path by the north pole and attracted by the south pole, and therefore it will trace out a curve extending from the north pole to the south pole of the magnet. Moreover, each point of this path will indicate *the direction of the resultant force,* and therefore *the curve at any point will be tangent to the force at that point.* A line of force, therefore, may be defined as *a contin-*

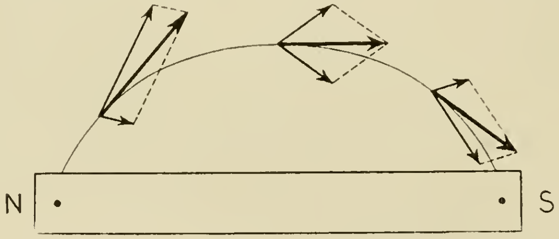

Fig. 227. A line of force is tangent to the direction of the force at every point.

uous line so drawn in a magnetic field that every point of that line is tangent to the direction of the force at that point. It is obvious, therefore, that *lines of force can never intersect;* for if they were thus to cross one another, the force would have two directions at the point of intersection, which is impossible.

Since the lines of force indicate the direction of the force about a magnet, these lines are thought also to indicate the directions along which the ether is strained in a magnetic field, and in our further study we may regard the lines of force and the strains in the ether as indentical in direction. It is carefully to be noted that the direction of a line of force at any pont in a magnetic field is defined as *the direction in which a small north magnetic pole would move,* or roughly, *the direction indicated by the north pole of a little magnetic needle* when placed at that point, and this direction should always be indicated by an arrow head.

357. Lines of Force and Lines of Magnetization. If by means of a little magnetic needle the field of force about a magnet is thus mapped out, it will be found that the lines of force all proceed from the north pole of the magnet and pass in continuous curves through the field and enter the south pole. These lines of force about a magnet are always closed curves, extending from the north pole to the south pole, the magnet itself closing the curve. Moreover, every line which emanates from the north pole comes back ultimately to the south pole, and flows back through the body of the magnet to the north pole, thus completing a closed circuit. Each line within the iron is continuous with a line of force in the space outside, Fig. 228; but within the iron these lines, which are due to the magnetized state of the

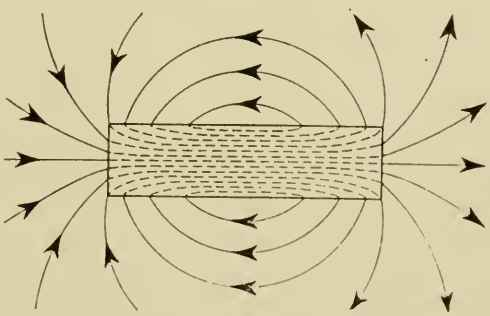

FIG. 228. The lines of magnetization complete and close the curves of the lines of force.

iron itself, have received the name of *lines of magnetization*. In a permanent magnet, therefore, lines of magnetization in the iron complete the curves of the lines of force in the air outside; and it is understood to be in this sense that lines of force always form closed curves.

A simple and effective method of mapping out the field of force about a magnet is to scatter iron filings upon a piece of glass or paper placed in the magnetic field. Since each little filing then becomes itself magnetic, as will be explained later, and so sets itself parallel to the lines of force, the filings arrange themselves in beautiful curves, Fig. 229, which show at once the character of the field of force about a magnet.

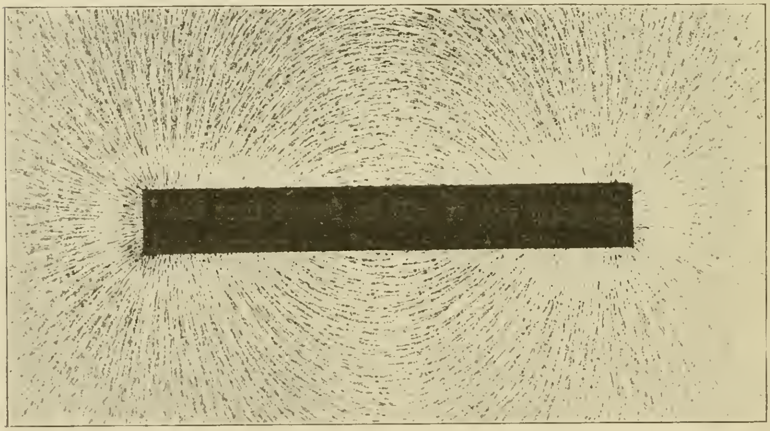

FIG. 229. The field of force about a magnet, mapped by iron filings.

358. Properties of Lines of Force. A careful study of the magnetic lines of force leads us to attribute to them the following properties:

1. They are without exception closed curves, the magnet, if they proceed from a magnet, closing the curve.
2. They never intersect.
3. They seem to act like tightly-stretched rubber cords.
4. They always tend to shorten themselves.
5. They seem to repel one another when parallel and in the same direction, and to attract and neutralize one another when parallel and in opposite directions.

359. Magnetic Phenomena Due to the Magnetic Field. If we conceive of lines of force as possessing the physical properties mentioned above, we may explain all magnetic phenomena as effects due to the lines of force with which we imagine a field of force to be filled. Thus, for instance, we can now understand why the north pole of one magnet attracts the south pole of another magnet. The lines of force emanating from the north pole of the first magnet enter at once the south pole of the second; and, as they leave the north pole of this magnet after passing through it, they pass in the closed curves of the lines of force around the two magnets, and so return to the south pole of the first, thus completing the curve. Because of that property of lines of force whereby they tend to shorten themselves, the two magnets, which are thus bound together by these lines of force passing through both, are now drawn closer together, just as they would be if tightly stretched rubber bands were thus to connect them.

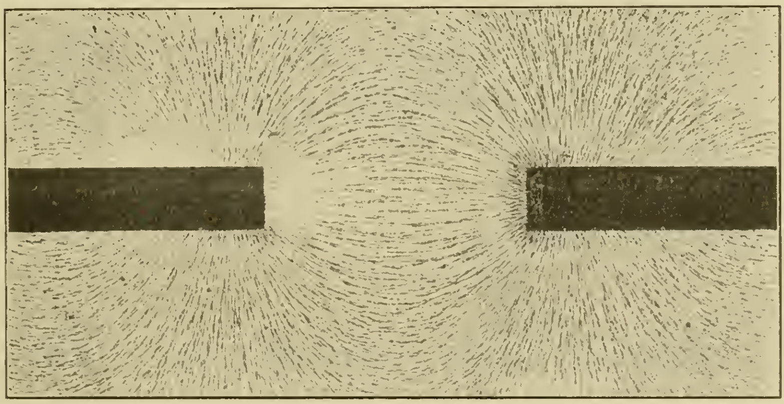

FIG. 230. Lines of force between unlike magnetic poles, producing attraction.

In an analogous way, two north poles or two south poles repel each other because the lines of force are crowded together in the space between the two poles; and as these lines cannot intersect or cross one another, the result is as if a quantity of elastic material were crowded between the two poles thus forcing them apart. The field of force, therefore, between two similar magnetic poles is strikingly different in character from that which exists between two dissimilar poles, as may beautifully shown by scattering iron filings upon a piece of glass placed over two magnets, arranged first with their unlike poles and then with their like poles toward each other, Figs. 230 and 231.

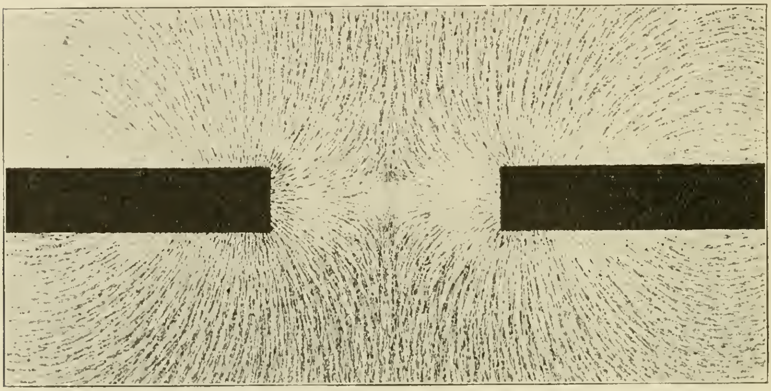

FIG. 231. Lines of force between like magnetic poles, producing repulsion.

360. The Intensity of the Field. The *strength* of the field, or the *intensity* of the field, is that characteristic of a field of force in virtue of which a magnetic pole is acted upon by a force when brought into the field. It is found that *the force which is exerted upon a magnetic pole in a field of force is proportional to the intensity of the field and to the strength of the pole.* That is, if m denotes the strength of the pole, and H the strength of the field into which this pole is brought then, if F is the force with which the field acts upon the pole,

$$F : Hm,$$

or

$$F = kHm,$$

where k is the factor of proportionality.

361. The Field of Unit Intensity. From the above equation it is possible to derive the definition of the field of unit strength: and in order that this may be expressed in C. G. S. units, F and m must likewise be expressed in these units, and k must equal unity. Accordingly, if we agree to call H unity when F and m are each equal to unity, then

$$k=1,$$

and the expression for the force F, acting on the pole m in the field of strength H, is

$$F=Hm.$$

In thus giving to k the value unity, we have, however, definitely determined that *the field of unit intensity is the field that acts with unit force upon the unit pole.* The C. G. S. unit strength of field, thus defined, is called the *gauss.*

By letting $m=1$ in the above equation,

$$F=H.\ 1,$$

from which it follows that *the strength of a magnetic field is measured by the number of dynes of force with which the field acts upon the unit pole.* Thus, for example, if a unit pole were acted upon by 1000 dynes of force when placed in a certain magnetic field, the field would then have a strength, or intensity, of 1000 units, or 1000 gausses.

362. Location of the Field of Unit Intensity. If the field of strength H is due to the presence of a pole of strength m, Fig. 232, we have, then, two expressions for the force acting on a magnetic pole m' which may be brought into this field:

$$F = \frac{m\ m'}{r^2}.$$

and

$$F=Hm'.$$

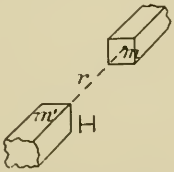

FIG. 232. Two expressions for the force acting on the pole m'.

The first of these relations gives the force that exists when the pole m' is brought near the pole m, and at the distance r from it (p. 409), while the second is the measure of the force exerted upon the pole of strength m' when this is brought into the field of strength H, due to the pole m (p. 427).

Equating these two equal expressions for F,

$$H\ m' = \frac{m\ m'}{r^2},$$

or

$$H = \frac{m}{r^2}.$$

Here H is the strength of the field at the distance r from the pole of strength m, which shows that *the intensity of the field due to a magnet pole varies inversely as the square of the distance from that pole.*

In the last equation, if $m=1$ and $r=1$ then

$$H = 1,$$

from which it follows that *the field of unit intensity is found at unit distance from the unit pole.*

363. Representation of a Magnetic Field. Since the intensity of a magnetic field is measured by the number of dynes of force with which the field acts upon the unit pole, we may *represent to the eye* the intensity of the field, Fig. 233, *by drawing just as many lines of force through each square centimeter of the field as there are dynes of force acting upon the unit pole.* That is, a field of 1000 units intensity would act with 1000 dynes upon the unit pole. This may be *represented to the eye* by drawing 1000 lines of force through a square centimeter, the plane of this little area being, of course, perpendicular to the lines of force.

Accordingly, the field of unit intensity would be represented by 1 line of force passing through each square centimeter of the field. But we have found that the unit field is found at unit distance from the unit pole. Therefore, *if we describe a sphere of 1 centimeter radius about the unit pole, the field of force everywhere on the surface of this sphere will be of unit intensity.* Moreover, the surface of this sphere will have an area of 4π square centimeters, because the surface of a sphere has an area of

$$4 \pi r^2$$

square centimeters, and in this case

$$r=1,$$

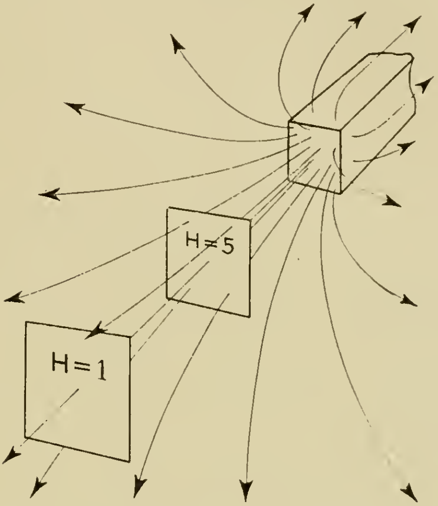

FIG. 233. The intensity of a magnetic field is represented by the number of lines of force passing through a square centimeter.

Accordingly, since the unit field about a unit pole is 4π square centimeters in extent, and since each square centimeter contains 1 line of force, *there are 4π lines of force which emanate from the unit pole. Therefore, a pole, the strength of which is m, sends out $4\pi m$ lines of force.*

364. A Uniform Magnetic Field. We have seen that from an isolated magnetic pole the lines of force diverge in all directions, and in the neighborhood of an actual magnet having two poles the lines of force pass in great curves from the north to the south pole. In such a field of force the direction as well as the intensity of the field is continually changing. If we could come into a region where the direction and the intensity of the field do not change, that is, where the lines of force are straight, equidistant, and parallel, Fig. 234, we should have a field of force which is uniform throughout. A *uniform* field of force, therefore, is a field in which the lines of force are *straight, equidistant, and parallel.* In the near neighborhood of poles or magnets, it is impossible to find a uniform field.

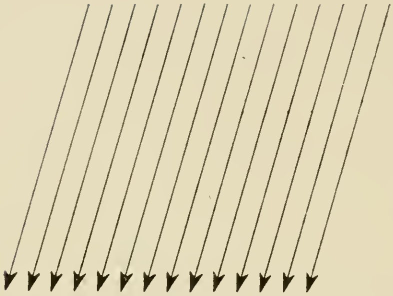

FIG. 234. A uniform magnetic field.

365. Moment of the Couple Tending to Turn a Magnetic Needle. When a magnetic needle is placed at right angles to the lines of force in a magnetic field, it is acted upon by a couple tending to turn the needle parallel to the lines of force of the field. If H is the intensity of the field and m the strength of pole of the needle, the force acting on each end of the needle is

$$Hm.$$

If the length of the needle is l, the moment of the couple tending to turn the needle is

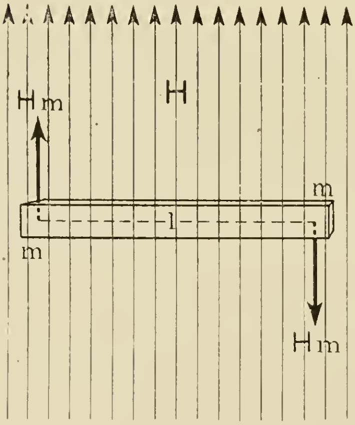

FIG. 235. A couple acts upon a magnetic needle.

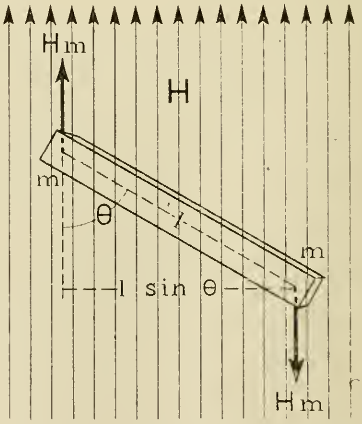

FIG. 236. A magnetic needle turns under the action of a couple.

$$Hml.$$

If the needle makes an angle θ with the lines of force of the field, the moment of the couple tending to turn the needle parallel to the lines of force is still the product of one of the forces and the perpendicular distance between them. The perpendicular distance between the forces in this case is

$$l \sin \theta,$$

therefore, the amount of the couple tending to turn the needle is

$$Hml \sin \theta.$$

If

$$\theta = 90°,$$

that is, if the magnetic needle is placed at right angles to the lines of force, this expression reduces to the one above.

366. The Magnetic Shell. A magnetic shell may be defined as a very great number of very short magnets arranged side by side like the cells in a honey-comb. The effect of this assemblage of short magnets would be to produce a magnetic sheet, one surface of which is composed entirely of the north poles of the little magnets, while the other surface is composed entirely of the south poles. When we attempt to magnetize a thin sheet of steel so that one entire surface is a great north pole while the other surface is a great south pole, we find that this arrangement is unstable, hence it is impossible to produce a magnetic shell in this sense.

As regards its external field of force, however, the magnetic shell, which can have no actual existence, may be replaced by a circular coil of wire through which a current of electricity is flowing. As will be shown later, the field of force about such a circular current is indentical with that which would be produced about the ideal magnetic shell. Accordingly, a circular current may in every instance replace a magnetic shell, as was done in the diagrams herewith given, and is therefore called *an equivalent magnetic shell.* The frequent occurrence of circular currents in the further study of electricity gives to the magnetic shell a place of great importance.

The *strength* of the shell is defined as *the magnetic moment per unit of surface,* that is, the strength of pole per unit of surface multiplied by the thickness, or the surface density of magnetization multiplied by the thickness.

367. The Mutual Action Between a Magnet and a Magnetic Shell. If we imagine ourselves to be looking at the edge of a magnetic shell, we shall then be able to map out the field of force about this shell, and this field will be found, Fig. 237, to be the same as that about a very short magnet with very broad poles. If the north pole of a magnet is now brought near the south pole of the magnetic shell, Fig. 238, the lines of force of the magnet will pass into the south pole of the magnetic shell, then out from the north pole of the shell around through the air to the south pole of the magnet, thus completing the circuit. The magnet and the magnetic shell are now bound together by the lines of force due to both, and these, as they tend to contract, draw the two together.

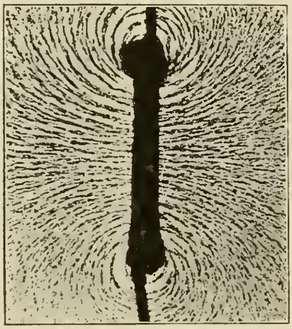

Fig. 237. The field of force about an equivalent magnetic shell.

It is important to observe that the magnetic shell will move in such a way that the greatest possible number of lines of force will pass through it, that is, in this case, the shell will move in such a way as to include not only its own lines of force but also those due to the magnet as well. If, however, the north pole of the magnet is approached to the north pole of the magnetic shell, Fig 239, the lines of force from these two sources will repel each other, and this will cause the magnetic shell and the magnet mutually to repel each other. In other words, the number of lines of force now passing through the shell is a minimum. Moreover, if the magnetic shell is free to move, it will move away from the magnet and turn around so as to present its south pole to the north pole of the magnet, and then attraction between the two will occur as before. In other words, if the like poles are together the magnetic shell cannot include the lines of force of the magnet because these are repelled by the lines of force of the shell. The shell, therefore, if it is free to move, will turn itself around so that its south pole will approach as near as possible to the north pole of the magnet, that is, so that the greatest possible number of lines of force may pass through it.

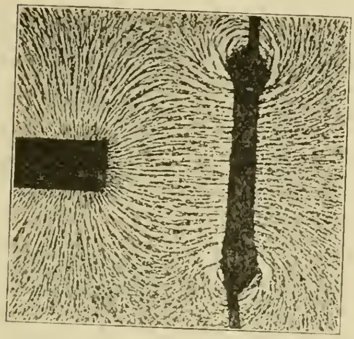

FIG. 238. The north pole of a magnet attracts the south pole of a magnetic shell.

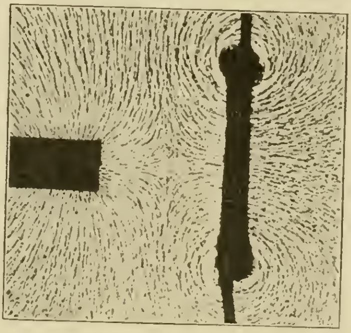

FIG. 239. The north pole of a magnet repels the north pole of a magnetic shell.

An extension of this principle to other cases gives rise to the following very general and very important law: *A magnetic shell in any magnetic field of force, if free to move, will set itself in that field in such a way as to include within its area the greatest possible number of lines of force.* Even this great law, however, is but a special case of the still greater law, previously mentioned (p. 122), which is of the very widest application in all branches of physics, namely: that if the parts of any system whatever are free to move, they will move in such a way that the resulting potential energy between these parts will be a minimum. In the present instance, the magnetic shell and the magnet possess a mutual potential energy so long as the two attract each other and are not in contact. As soon, however, as they move, the mutual potential energy which they possessed in their separated positions becomes kinetic, and they are able to do work in virtue of the motion with respect to each other. When one pole of the magnetic shell is in contact with the opposite pole of the magnet, no further motion is possible, and therefore the store of potential energy beteween these two is reduced to zero, and the two are in the position of maximum stability. In obedience to the general law, accordingly, *the parts of this system move in such a way that the potential energy between the two becomes a minimum.*

CHAPTER XXIX.

MAGNETIC INDUCTION.

368. Magnetic Induction. In a previous chapter (p. 410), we have seen reason to believe that an unmagnetized bar of iron or steel consists of molecules, each completely magnetized and having a north and a south pole of its own, but forming molecular groups among themselves in obedience to their mutual polar attractions. The process of magnetization consists, then, not in magnetizing the molecules, for these are already magnetized, but in destroying these molecular groups, and compelling the molecules to assume an orderly direction.

We understand now that if each molecule is a complete magnet, it must be surrounded by a tiny molecular field of its own; and as the molecules arrange themselves into groups, these molecular magnetic fields must have all possible directions, and thus produce no resultant effect. If, however, we bring another magnet, or a magnetic field from some other source, near this unmagnetized bar of iron, we introduce into the latter the magnetic field which we thus bring up. The presence of the magnetic field which we thus introduce into this bar of iron compels the molecules of the bar to set themselves parallel to the lines of force of this field. In other words, the presence in the bar of this extraneous field, causes the molecules of the bar to arrange themselves so that their own molecular fields coincide in direction, and produce a resultant field in the iron parallel to the lines of force of the field which is thus introduced.

By this process, which is known as *induction,* the bar of iron becomes magnetized simply by bringing it into the presence of a magnetic field from some other source. In all instances of the attraction of iron by a magnet, the magnetization of the iron by induction must precede the attraction.

369. The Permeability of Iron. When a bar of iron is magnetized by induction, it is obvious that there are within the iron two sets of lines of force, one due to the outside field, which thus compelled the iron to become magnetic, and the other due to the lines of magnetization of the iron (p. 423) from the fact that the iron itself is now a magnet. *These two independent sets of lines of force, taken collectively, represent the total number of lines of force within the iron, and are known as the lines of induction.* Accordingly, the number of lines of induction which pass through the iron is enormously greater than is the number of lines of force from the outside field which would have passed through the same space, if the latter had been filled with air instead of iron. This is expressed by saying that the iron has greater *permeability* than air, where permeability is defined as the ratio of the number of lines of induction per square centimeter which pass through the iron to the number of lines of force per square centimeter which would pass through the air, if the same space were filled with air instead of iron. Thus, if H denotes the intensity of the field in gausses (p. 427), that is, the number of lines of force per square centimeter when a certain space is filled with air, and B denotes the number of lines of induction per square centimeter when this space is filled with iron, the ratio of B to H denotes the permeability μ. Accordingly,

$$\mu = \frac{B}{H}.$$

The permeability of iron varies with the quality of the iron and with the degree of magnetization. For a good quality of wrought iron, a magnetizing force that will develop 50 lines of force per square centimeter in a space filled with air will develop about 16,000 lines of induction per square centimeter if this space is filled with iron. In this case μ is about 320.

370. Conductivity of Iron for Lines of Force. A piece of iron which is magnetized by induction becomes, as we have seen, a complete magnet with an independent magnetic field of its own. When, however, we bring a piece of soft iron into the magnetic field of a permanent magnet, and scatter iron filings upon a piece of glass which may be laid over both, we find two surprising facts from a study of the iron-filing pattern produced. First, that the soft iron does not have about it the normal magnetic field which we have learned (p. 424) exists about a magnet; and, second, that the field of force of the permanent magnet is strangely distorted, the lines of force apparently being deviated so as to flow into the soft iron at one end and out at the other, having thus passed entirely through the iron, as may be seen in Fig. 240. This phenomenon has led to the belief that the soft iron is a good conductor for lines of force, and that these lines will thus deviate from their normal course in order to pass through the iron, which, for them, is a better conducting medium than air.

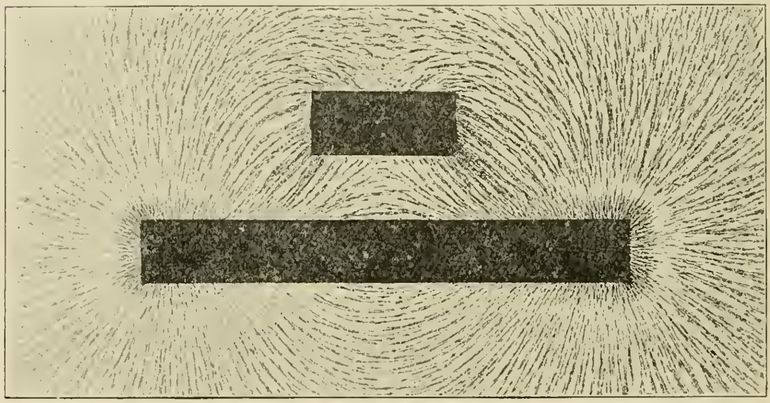

FIG. 240. A piece of soft iron is apparently a good conductor for lines of force.

A few minutes' reflection, however, will show that this effect is only apparent and is really due to another cause, although its conception is so convenient and helpful that we constantly refer to the idea of the great conductivity of iron for lines of force. In the crude diagram, Fig. 241, the curved lines *A, B, C, D, E,* represent the lines of force in the normal magnetic field about the permanent magnet *N S,* while *N′ S′* represents a piece of soft iron in the field of this magnet. Certain of the lines of force from *N S,* represented by *B* in the diagram, will pass through the soft iron, thus causing the latter to become a magnet by induction. The end of the soft iron into which these lines of force pass becomes a south pole, and the end from which they leave becomes a north pole; so that the soft iron becomes a complete magnet with poles *N′ S′* of its own, and its own magnetic field, represented by the lines of force *F* and *G.*

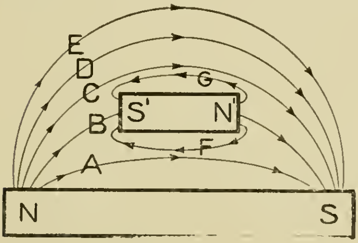

FIG. 241. A piece of soft iron magnetized by induction.

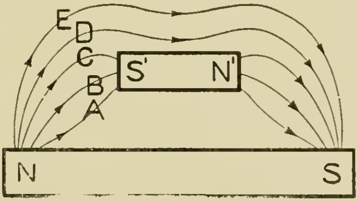

FIG. 242. The resultant field, showing apparent conductivity of soft iron for lines of force.

Such, indeed, would be the two fields of force, which actually exist, were it not for the fact that certain properties of lines of force necessitate the union of several separate fields into one resultant field. We have learned (p. 424) that lines of force which are parallel and in the same direction will repel one another laterally, while lines of force which are parallel and in opposite directions attract and neutralize one another. Both of these properties of lines of force find illustration in our diagrams. In Fig. 241, and in every diagram of a magnetic field, the lines of force, represented by *A, B, C, D, E,* which are of necessity near together as they enter and leave the magnet, are much farther apart opposite the central portion of the magnet, due to the mutual repulsion of lines of force which are parallel and in the same direction. Moreover, the lines *A* and *F,* and also the lines *C* and *G,* being parallel and in opposite directions for the central portions of their paths, attract and neutralize each other, and therefore disappear from this part of the field. Furthermore, as lines of force must always be closed curves, and as they always tend to contract, the ends of the lines *A* and *F,* and also the ends of the lines *C* and *G,* unite when their central portions disappear, thus giving the resultant field, shown roughly in Fig. 242.

From this it is evident that the lines of force from the north pole of the magnet *N S* deviate from their normal path and crowd in great numbers into the nearer end of the soft iron. It is likewise evident that these lines of force pass, apparently with great freedom, through the iron, and emerge from the remote end to continue their course through the air to the magnet's south pole. Indeed, the soft iron does actually carry not only the lines of force of the field in which it is placed, but also its own lines of force, due to the fact that it itself has now become a magnet by induction. Though the effect is only apparent, it is, however, correct in every way to speak, in this sense, of soft iron as being a good conductor of lines of force, and of lines of force in a magnetic field as deviating from their normal path in order to pass through a piece of soft iron that may be present in the field.

The central portion of the line *C* having disappeared, this line no longer repels the line *D*, so that the central portion of the latter line bends down towards the magnet, being repelled by *E*, while this line is likewise bent downward by the repulsion of lines still farther above, and so on.

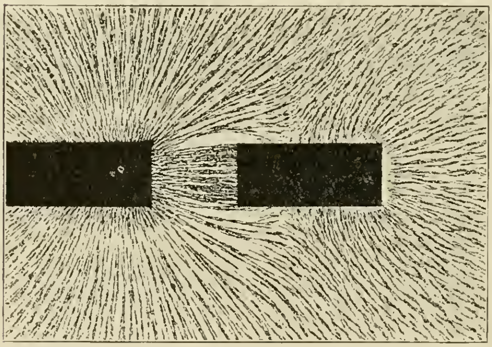

FIG. 243. Magnetic induction—Apparent conductivity of soft iron for lines of force.

A similar instance of the distortion of the field, and of the apparent conductivity of soft iron for lines of force, may be seen in Fig. 243, where the soft iron is placed near the end of the permanent magnet.

371. Magnetic Shielding. The conclusions reached in the last paragraph show that the field of an induced magnet to some extent neturalizes and therefore weakens portions of the magnetic field in which it is placed. Advantage has been taken of this weakening of the field, due to the presence of the soft iron, to produce the effect known as *magnetic shielding.* Since the results of this weakening are in every way the same *as if* the iron were a better conductor than air for lines of force, we endeavor to surround by soft iron the region to be shielded, so that the lines of force from an external magnetic field will pass through the iron rather than through the shielded region.

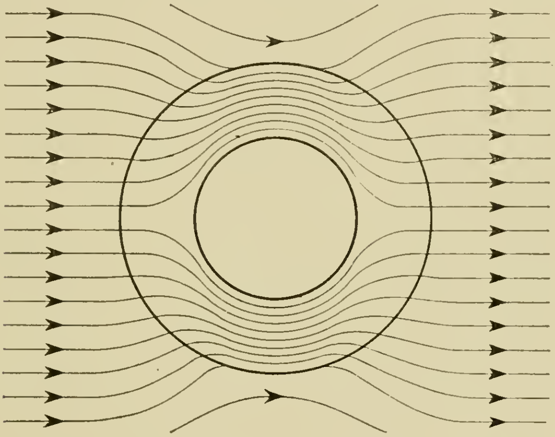

FIG. 244. The principle of magnetic shielding.

Thus, Fig. 244, if a hollow cylinder of soft iron is placed in a magnetic field, the lines of force of the field will enter the iron on one side, and after passing through the iron will leave from the other, thus almost completely shielding the inner region from the presence of lines of force. For the same reason, watches, which have to be carried near dynamos or strong electromagnets, and which would be injured if magnetized, are provided with so-called magnetic shields, which are simply heavy cases of soft iron.

CHAPTER XXX.

THE EARTH'S MAGNETIC FIELD.

372. The Earth's Magnetic Field. Whenever a magnet is suspended so as to turn freely in a horizontal plane near the surface of the earth, it is acted upon by a force which causes it to point north and south. In other words, there is a magnetic field of force all about the earth, the lines of force of this field having a general north and south direction. The earth itself, therefore, must be regarded as a great magnet having its south pole located somewhere in the northern hemisphere and its north pole located somewhere in the southern hemisphere. This is the inference that must be drawn as to the nature of the poles of the earth if we assume, as is done in all English-speaking countries, that the north pole of a magnetic needle is the end which points north, and that unlike poles attract each other. The lines of force in the earth's magnetic field proceed, therefore, from the earth's magnetic north pole in the southern hemisphere, and pass in great curves of lines of force about the earth and enter the magnetic south pole in the northern hemisphere.

373. The Dip or Inclination of the Needle. We can ascertain the direction of the lines of force at any point in the earth's magnetic field by employing the device known as the dipping needle. The dipping needle consists of a magnetic needle mounted so as to turn freely in every direction, that is, so as to move in a vertical as well as in a horizontal plane. We know that a magnetic needle in a field of force will set itself parallel to the lines of force, and will lie in such a direction that the lines of force of the field will enter the south pole and flow out from the north pole, after passing the whole length of the needle. When such a needle is mounted in the earth's magnetic field in the latitude of Madison it is found that the north pole dips downward, the needle itself making an angle of about 74° with the horizontal, from which we conclude, that the lines of force of the earth's field at this point are not horizontal but are inclined at an angle of 74° to the surface of the earth. This angle which the dipping needle makes with the horizontal, that is, the angle at which the lines of force of the earth's field are inclined to the surface of the earth, is called the *dip, or inclination,* of the needle.

It is, of course, obvious that as the dipping needle is carried farther to the north, the angle of dip will increase, and as it is carried farther to the south, this angle will decrease. It is the dipping needle, therefore, that enables us to determine the location of the magnetic poles of the earth, because at the magnetic poles the lines of force are vertical. It remains, therefore, simply to carry the dipping needle over the surface of the earth until a point is found where the angle of dip is 90°, that is, where the dipping needle stands vertically. In this way Sir John Ross, in 1831, found the north magnetic pole to be, not at the geographical north pole, but in Lat. 70° 5′ N., Long. 96° 33′ W., that is, at a point between Hudson's Bay and Baffin's Bay. Strangely enough, the south magnetic pole, as found by Sir Ernest Shackleton in 1909, is not at a point diametrically through the earth, but is located in South Victoria Land, Lat. 72° 25′ S., Long. 154° E., only 110° of longitude west of the north magnetic pole. The magnetic equator is the somewhat irregular line extending around the earth, sometimes in the northern hemisphere and sometimes in the southern, where the angle of dip is zero, that is, where the dipping needle lies horizontal.

The earth's lines of force, as we detect them near the surface, are really portions of great curves extending from the magnetic north pole in the southern hemisphere to the magnetic south pole north of Hudson's Bay. For a limited region, however, such as a room or a laboratory, these lines of force, which are very short lengths of very great curves, may be considered as straight, equidistant, and parallel, and therefore produce a uniform field (p. 430). The earth's field is thus regarded as a uniform field of force, although in modern steel-constructed buildings, with pipes of iron for gas, water, and steam, the uniform field of the earth is sadly distorted.

Measurements of the intensity of the earth's field show that a unit pole placed in this field would be acted upon by about half a dyne of force. In other words, we may think of one line of force passing through about every two square centimeters of area, this area being taken perpendicular to the direction of the lines of force. The earth's field, therefore, is relatively a very weak field.

374. The Horizontal Intensity of the Earth's Magnetic Field. If we represent by a line the intensity as well as the direction I of the earth's magnetic field, we may resolve this total intensity into two components, Fig. 245, one vertical and the other horizontal. The horizontal intensity of the earth's magnetic field, therefore, is the component of the total intensity of the earth's field which lies in a horizontal direction. It is only the horizontal component of the earth's magnetism that is effective in causing the magnetic needle to point north and south, and hence this is the element with which we most frequently have to deal. The vertical component tends to cause the needle to dip; but in the ordinary magnetic needle, which can move only in a horizontal plane, this dipping motion is prevented by the action of the pivot upon which the needle turns. The horizontal intensity is always denoted by H.

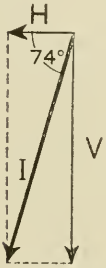

FIG. 245. The horizontal intensity.

375. The Declination of the Magnetic Needle. The Magnetic meridian at any place on the earth's surface is the vertical plane at that place lying in such a direction as to contain the axis of the magnetic needle. Accordingly, the magnetic meridian at any place is the vertical plane parallel to the lines of force at that place.

Since the earth's magnetic pole located in the northern hemisphere is not at the geographical north pole of the earth, the magnetic needles all over the earth, in pointing to the magnetic pole, point more or less exactly to this region north of Hudson's Bay and not to the true geographical pole. At any place, the deviation which the magnetic needle makes with the true north and south direction, that is, the deviation of the magnetic meridian at any place from the geographical meridian, is called the declination of the needle.

At certain places on the earth, the magnetic needle, in pointing to the magnetic pole of the earth, points also to the geographical pole, and a line connecting these places is called the *agonic line, or line of no declination.* Such a line extends southward from the magnetic pole, and after running in a somewhat irregular course through Michigan, Ohio, and Kentucky, finally passes out into the Atlantic Ocean not far from Charleston, South Carolina. Everywhere along this line the magnetic meridian coincides with the geographical meridian, and the magnetic needle points north. To the east of this line the magnetic needle, in pointing to the magnetic pole of the earth, points west of the true north and south direction, and to the west of this line, for the same reason, the magnetic needle points somewhat to the east of the north and south direction. For all places, therefore, east of the agonic line the declination is west, and, similarly, for all places west of this line the declination is east. Across the United States the declination varies from about 24° west in the extreme eastern part of Maine to 23° east on the coast of the state of Washington.

The dip of the needle, the declination of the needle, and the intensity of the earth's field, are called *the three elements* of the earth's magnetism. Inasmuch as the total intensity of the earth's field is difficult to determine directly, this is usually calculated from the angle of dip and the horizontal intensity, since both of these quantities admit of ready measurement.

376. Changes in the Earth's Magnetic Condition. One remarkable phenomenon in connection with the earth's magnetism is the fact that all of the elements are undergoing change. Even the magnetic pole and the line of no declination are slowly shifting. To these changes there seems to be a daily period and also a period lasting twenty-eight days. Moreover, the magnetic conditions very nearly repeat themselves every year, and again every eleven years, so that there is a yearly period and a period lasting eleven years, as well as a cycle which requires centuries to complete. In addition to these periodic variations there are frequent and sudden irregular and violent changes which are called magnetic storms.

While nothing is known as to the origin of the earth's magnetism or of the changes which this is continually undergoing, the remarkable fact remains that these periodic variations, as well as the irregular disturbances, or magnetic storms agree precisely in point of time with occurrences taking place upon the sun. The twenty-eight-day period of magnetic changes coincides with the time of rotation of the sun upon its axis, and the eleven-year period is the same as the eleven-year period of the recurrence of sun spots. Magnetic storms invariably occur after an unusual outburst upon the sun, or the sudden appearance of a large spotted region. Moreover, these irregular and violent magnetic disturbances are almost always accompanied by an unusual display of the aurora borealis, or northern lights.

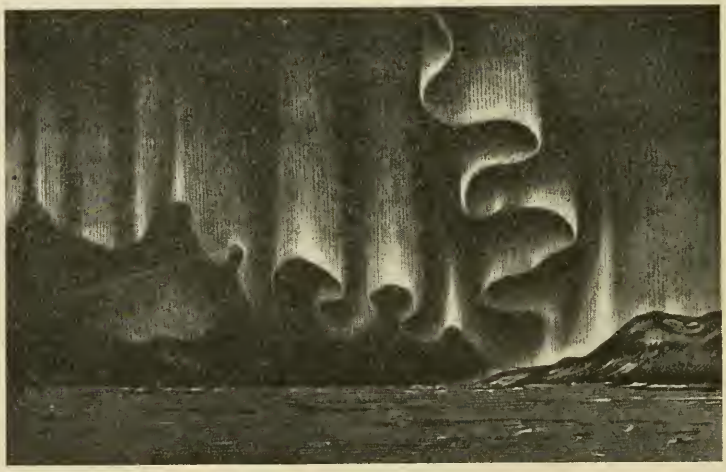

Fig. 246. The Antarctic auroral curtain—By permission of the Royal Society.

CHAPTER XXXI.

DIAMAGNETISM.

377. Diamagnetism. As regards their magnetic properties, all substances may be divided into two classes. Bodies like iron, nickel, cobalt, chromium, and many others, which are attracted by a magnet, and which tend to move from the weaker into the stronger parts of a field which is not uniform, setting themselves parallel to the lines of force of this field, are called paramagnetic, or simply magnetic. Other bodies, of which bismuth is the great type, which are apparently repelled by a magnet, and which tend to move from the stronger into the weaker parts of a field which is not uniform, thus setting themselves at right angles to the lines of force of this field, are called diamagnetic. If a bar of bismuth is suspended by a silk fiber in the intense magnetic field between the poles of a large magnet, it will be seen to set itself at right angles to the lines of force of the field. This phenomenon was once thought to be due to the fact that bismuth is repelled by a magnet, but later investigation has shown that this effect is undoubtedly due to another cause.

378. Theory of Diamagnetism. When we remember that lines of force flow from the north pole of a magnet to the south pole ordinarily through the ether, and that iron offers a far better conductor for lines of force, we may also conceive of there being other substances through which lines of force will flow with greater difficulty than through the ether. If we then assume that the ether is a conductor for lines of force which is ever present and in which all other bodies are immersed, we see that whenever a substance is placed between the poles of a magnet the lines of force have the choice of two paths, one through the body and the other through the surrounding ether. If the body is like iron and proves to be a better conductor for lines of force than the ether, the substance will, so to speak, crowd the ether out of the way, and set itself with its longest axis parallel to the lines of force, so that the lines of force may pass as far as possible through it. Substances like iron, therefore, which conduct lines of force better than the ether in which they are placed are called magnetic. When, however, a substance like bismuth sets itself with its longest axis at right angles to the lines of force, we believe that this is due to the fact that the substance is a poorer conductor for lines of force than the surrounding ether, and that the ether, as it were, crowds this poorer conductor out of the way, so that the lines of forces may pass the smallest possible distance in it. Substances, therefore, like bismuth, which do not offer so ready a path for lines of force as the ether, are called diamagnetic. In other words, the permeability (p. 435) of a diamagnetic substance is less than that of air, which is taken as unity. The relative conductivity for lines of force of iron and of bismuth is shown in the accompanying diagrams, Fig. 247 and 248.

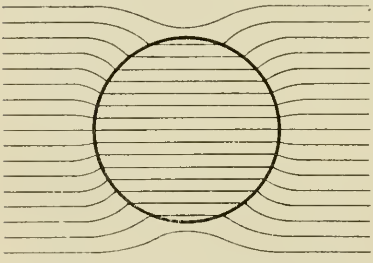

FIG. 247. A sphere of soft iron in a uniform magnetic field.

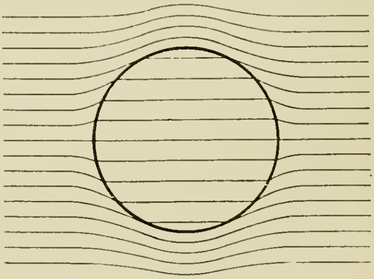

FIG. 248. A sphere of bismuth in a uniform magnetic field.

379. Diamagnetic Phenomena. This theory of diamagnetism may be beautifully illustrated by the following experiment. If a small tube containing a solution of sulphate of iron is suspended horizontally between the poles of a powerful magnet, the tube will set itself parallel to the lines of force, showing that by itself it is magnetic. If, however, this is immersed in a tank containing a solution of the chloride of iron, which is more magnetic than the sulphate, the little tube will be crowded out of its former position by the more magnetic chloride of iron, and will set itself at right angles to the lines of force, showing that the magnetic sulphate of iron behaves like a diamagnetic body when it is immersed in a fluid more magnetic than itself.

If a soap-bubble is filled with oxygen gas and placed between the poles of a magnet, the bubble will be elongated in the direction of the lines of force, showing that oxygen is magnetic. If the same experiment is tried with hydrogen, the bubble will be flattened slightly on the sides nearest the poles, proving that hydrogen is diamagnetic. A better way to show his latter fact is to send a stream of hydrogen, made visible by a white cloud of ammonium chloride, between the poles of the magnet. The hydrogen, being a poorer conductor of lines of force than the ether, will be crowded aside and turned out of the field. A candle flame, when placed between the poles of a magnet, is seen to be depressed as if blown by a current of air. This shows that a flame is diamagnetic and is pushed aside by the better conducting medium around it.

Particular mention must be made of the fact that all diamagnetic phenomena are exceedingly feeble as compared with those manifested by paramagnetic substances.

CURRENT ELECTRICITY.

CHAPTER XXXII.

GALVANIC BATTERIES.

380. The Need of Maintaining a Difference of Potential. In the preceding pages we have seen that when two bodies at different potentials are connected by a conductor, electrons flow from the one at lower potential towards the one at higher potential. We have also seen that this flow, which continues as long as there is a difference of potential, very quickly ceases, inasmuch as the two bodies soon become at the same potential. In the early days of the science, when electricity was though to be a fluid, the flow or passage from one conductor to another was called the electric current; and in modern times, although the idea of a fluid is no longer held, the word current is still retained, and the current is said *to flow* through a conductor. Since electricity at rest, or in the static condition, can do no work, it is only during the flow of the current that energy from this source can be obtained; hence it is of the utmost importance to investigate the conditions under which a difference of potential, after being created, may be maintained at a constant value, even while the current flows.

When a conductor unites the two coatings of a Leyden jar, or when connection is made between a charged body and the earth, a flow of electrons takes place, and this, in every sense of the word, is a current. This flow is but momentary, or consists at most of but a few surges back and forth, and these soon cease. It is true that the continuous action of the Holtz machine maintains one of the conductors at high potential and the other at low potential, so that if these are connected, electrons flow along the path thus provided. But since the machine is largely of glass, and the electrons in their redistribution must flow from points off upon the glass plate and from the plate again to the points, the internal opposition to the flow is so great that the relative number of electrons which this or any similar device is capable thus of setting in motion is really very small. Moreover, the potentials which a machine of this kind creates are enormously great, which makes it quite impossible sufficiently to insulate a conductor so that electrons will not escape to neighboring bodies, and, by ionizing the adjacent air, to a great extent destroy the potentials which have been produced. The discovery, therefore, made by Volta, that a difference of potential may not only be *produced,* but *maintained* for long periods in a far simpler and more effectual way, has proved to be of great value to the world.

381. Volta's Discovery. In the closing years of the eighteenth century the Italian physicist, Volta, found that a very appreciable difference of potential may be produced by placing a disc of zinc upon a disc of copper, then a disc of cloth moistened with salt water upon the zinc, then another disc of copper, and upon it another disc of zinc, with another piece of cloth moistened as before upon the zinc, and so on in this order, until a pile of considerable height is obtained. Indeed, the gold-leaf electroscope may be charged simply by touching it to one end of the pile so formed. Moreover, if the two ends of the pile are now connected, electrons will be found to flow *continuously* through the conductor, showing not only that a difference of potential is *produced* by the pile, but that the action of the pile is such that this difference of potential is *maintained.* Soon after this discovery, Galvani, a pupil of Volta, showed that the essentials for the production and maintenance of this difference of potential are simply two dissimilar metals and a fluid that acts chemically more violently upon one than upon the other. It is in the discoveries of these two men that the galvanic battery cell of to-day finds its origin.

382. The Production of a Difference of Potential. If a strip of copper and a strip of zinc are dipped into a dilute acid, such as sulphuric acid, they are found at once to be at slightly different potentials, due to the difference in the chemical action of the acid upon the two metals. The differences of potential, however, which can thus be produced by chemical action are so small as compared with those which we have heretofore considered, that a means of detecting them has been designed which is far more sensitive than any of the electroscopes previously mentioned. This wonderfully delicate instrument, which not only detects differences of potentials but measures them, is known as the quadrant electrometer, and was devised by Sir William Thomson, the late Lord Kelvin.

383. The Quadrant Electrometer. The essential features of the quadrant electrometer consist of certain fixed parts called the quadrants, and of a movable part called the needle. The quadrants, Fig. 249, are the four quarters of a round, flat, hollow brass box, slightly separated from one another and mounted upon glass rods for insulation. The opposite diagonal quadrants are connected by a fine wire, and so always remain at the same potential. The "needle" consists of a piece of thin aluminum cut into a shape somewhat resembling a dumb-bell, or the figure 8. This needle is hung by a very delicate thread from the upper part of the case, and swings in a horizontal plane exactly in the center of the four quadrants. A wire hangs vertically downward from the center of the needle and dips into a cup of strong sulphuric acid beneath. A charge of electricity which may be given to this acid is conducted through the wire to the needle above, so that the needle is charged as it swings to the right and left in the center of the quadrants,

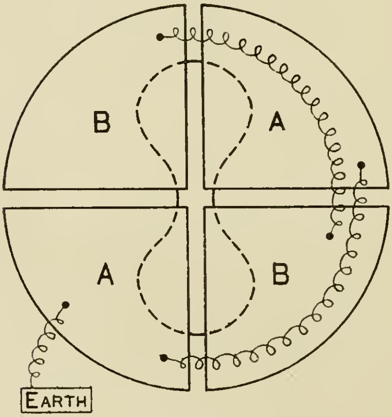

FIG. 249. The quadrant electrometer.

turning on its suspending fiber as an axis. Indeed, the acid in this cup really forms the inner coating of a Leyden jar and thus serves the double purpose of charging the needle from the store of electricity with which it itself is charged, and also of keeping the interior of the instrument dry; for the sulphuric acid very readily absorbs the moisture which would otherwise collect on the insulating glass supports of the quadrants, and thus make them partial conductors. A small mirror, which is rigidly attached to the needle and turns with it, reflects a beam of light, so that the motion of the spot of light serves as an indication of the movement of the needle.

If both pairs of quadrants are connected with the ground, their potentials are zero, and the needle is neither attracted by one pair nor repelled by the other, and therefore remains at zero. If, however, one pair, as, for instance, A, remains connected with the ground while the other pair, B, is connected with an electrified body as a source, electrons will flow either to these quadrants or away from them, thus giving them the potential of the source. The quadrants B, now being electrically charged, will repel or attract the needle according as the charge upon the quadrants is of the same or opposite sign as that upon the needle. If this pair of quadrants is now connected with a body of opposite electrification, the quadrants will acquire an opposite charge and the needle will turn in the opposite direction. The direction of the movement of the needle, as shown by the direction of the motion of the spot of light, indicates the nature of the charge and therefore also the potential of the source, since we now assume the source to be all alone by itself and uninfluenced by other electrified bodies. The needle has been given its peculiar shape in order that the amount of the deflections may be proportional to the magnitude of the potentials which produce them.

384. The Potentials in a Galvanic Battery. It is, therefore, with the quadrant electrometer that we are enabled to detect and measure the potentials which are produced when a strip of copper and a strip of zinc are placed near each other, but not in contact, in a vessel of dilute acid. In virtue of the chemical action, which is greater upon the zinc than upon the copper, these two metals become at different potentials as soon as they are immersed in the liquid. If the strip of zinc is connected with the earth and the strip of copper joined to the quadrants B of the quadrant electrometer, the quadrants A being likewise connected with the earth, the needle will move in such a direction as to indicate that the copper is at high potential, that is that electrons have flowed from the quadrants B to the copper. If the copper is connected with the earth, and the zinc connected with the quadrants B, the needle will turn in the opposite direction, indicating that the zinc is at low potential, or that electrons have flowed from the zinc to the quadrants. Here the difference of potential is brought about not by means of mechanical work performed, but because of the difference in the chemical energy expended by the action of the acid upon the two metals. In this case, therefore, it is chemical action that is the agent producing the difference of potential; and as long as this chemical action continues, so long will the copper remain at a higher potential than the zinc.

Accordingly, if these two plates are connected by a wire, electrons will flow from the zinc through the wire to the copper, tending to raise the potential of the zinc and lower the potential of the copper. *But chemical action continues,* so that a difference of potential *is maintained* between the zinc and the copper, and hence *a constant stream of electrons flows from the low potential zinc to the high potential copper,* and an electric current is said to flow through the wire.

Any device, capable of developing and maintaining a continuous difference of potential by virtue of chemical action, is called a galvanic battery, or sometimes a voltaic battery, or a voltaic cell. The differences of potential that can be developed by chemical means are very small as compared with those which we have studied under static electricity.

385. The Direction of the Current. In a previous section (p. 350) it has been pointed out that it would be perfectly natural and most reasonable to speak of the direction of the current as the direction in which the electrons flow. This indeed would be the case if we were at this time forming the electrical phraseology. Ever since the time of Franklin, however, a positively electrified body was conceived to be one which had upon it an excess of electricity, and it was thought to be this abnormal quantity, or this positive charge, that flowed away when the body was discharged. Moreover, by analogy, it was believed in all instances to be the positive charge that moved when a current from any source flowed through a conductor.

When it was found, therefore, that a flow of current could be obtained most effectively and efficiently from a battery, it was assumed here also that a positive charge moved, and a current was said to flow through the conductor from the high potential copper to the low potential zinc. This, through many years of use, has come to be recognized as the direction in which the current flows, and is still so considered, although we now know that the flow of electrons through the conductor takes place from the low potential zinc to the high potential copper. Accordingly, out of deference to long-established scientific, as well as popular, usage, we are constrained to speak of the *current* as flowing through the conductor from the copper to the zinc, although we are perfectly well aware that the flow of *electrons* which constitutes the current takes place in the opposite direction.

386. Electromotive Force. The name *electromotive force* has been given to the cause of whatever nature that is capable of producing a difference of potential. An electromotive force therefore may be defined as anything that will produce a difference of potential. In the case of the glass rod and the piece of silk, for instance, the electromotive force is the work that is done in separating the two after electrification; in the case of the electrophorus, the electromotive force is the work that is done in lifting the cover from the negatively charged plate; in the case of the Holtz machine, the electromotive force is the work done in turning the plate against the electrical forces developed; in the case of the galvanic battery, the electromotive force is the chemical action. The word electromotive force is usually written E. M. F., and later we shall shorten it still further to E. Inasmuch as any cause which produces a difference of potential is an electromotive force, these two quantities are measured by the same units and are frequently used interchangeably. Electromotive force is unfortunately named, since it is not force but energy, and may, like potential, be expressed in terms of energy expended or work performed.

387. Dissociation in Dilute Aqueous Solutions. In treating of osmotic pressure (p. 230), it was shown that when a solution of a substance like sugar is separated from pure water by a porous membrane, the water tends to pass through the pores of the membrane and mingle with the sugar solution. Moreover, the tendency of these two liquids thus to mingle is so great that the water will pass through the pores of the membrane and mingle with the solution within, even though in its passage it causes a pressure upon the membrane amounting to several hundred atmospheres. This pressure, which the tendency of the two liquids to mingle is capable of producing, or, in other words, the counter pressure which it would be necessary to exert upon the other side of the membrane to prevent the water from passing through, is called the osmotic pressure.

It was furthermore shown that the osmotic pressure which is thus capable of being produced is directly proportional to the number of molecules of the dissolved substance present. If, therefore, a certain number of molecules of the sugar are dissolved, a certain osmotic pressure will be produced, but if the number of the dissolved molecules is doubled, the osmotic pressure will be doubled. In addition to this, the remarkable fact was brought out that while the osmotic pressure depends upon the number of molecules in solution, it is apparently independent of the nature of the molecules, that is, it makes no difference what kind of substance is dissolved, so long as the same number of molecules are present. If, therefore, the same number of dissolved particles of any other substance is employed, the osmotic pressure will be the same as it is when sugar is used.

When, however, we dissolve a given number of molecules of sugar in a certain volume of water and determine the osmotic pressure of the solution, and then dissolve the same number of molecules of common salt, NaCl, in an equal volume of water and determine its osmotic pressure, we shall find this pressure to be just twice as great in the latter case as in the former. In view of the facts which we know to be true regarding osmotic pressure, the evident explanation of this phenomenon lies in the inference that the solution of NaCl must contain twice as many particles as does the solution of sugar. We are thus forced to the conclusion that every molecule of NaCl as it dissolves, breaks into two parts; and accordingly, since there are twice as many particles in the solution, the osmotic pressure is twice as great.

Inasmuch as the molecule of NaCl consists of one atom of Na and one atom of Cl, the only way in which this molecule can break into two parts is for the sodium to separate from the chlorine; and so we believe that the Na particle floats away by itself and the Cl particle does likewise. The breaking up of the molecule of a compound substance when this is dissolved in water is called *dissociation*, and this dissociation is more complete the more dilute the solution. *The component parts into which a compound substance breaks up by dissociation are called the ions.*

388. The Nature and Importance of the Ions. One fact of the most momentous import has been observed regarding the conductivity of solutions for electricity. This one fact, which perhaps more than any other has revolutionized our whole view regarding the generation of a difference of potential in the galvanic battery, and which alone has compelled physicists to formulate an entirely new theory regarding the chemical relations of the electric current, is this: that *only those solutions which dissociate and form ions when dissolved conduct electricity.* A solution of sugar, therefore, which does not dissociate when dissolved does not form ions, and so does not conduct electricity; on the other hand, a solution of common salt, which does dissociate and does form ions, is found to be a good conductor. This remarkable discovery points undoubtedly to a very close relation between the ions of a dissociated substance and the conduction of electricity by a liquid.

At first, the phenomenon of dissociation seems to be wholly at variance with the ideas which we have formerly had regarding physical and chemical theories. We have learned (p 8) that an atom cannot exist by itself; and chemistry teaches that a particle of free sodium cannot remain in water because a chemical action takes place between the two.

We must remember, however, that whenever any two dissimilar substances are separated, the electrons which the two bodies possess together are redistributed (p. 342), one carrying with it electrons in excess of its normal number, and the other being left with a deficit equally great. In other words, one body is charged negatively due to the additional electrons which it carries, and the other is left charged positively to the same degree because of the electrons which have in this way been taken from it.

This law, to which there are no exceptions, is found likewise to hold true when the Cl and the Na separate from each other by dissociation. When this separation is effected and ions are thereby produced, the Cl is found to carry with it one electron in excess of its usual number, and is therefore charged negatively, while the Na is left with one electron less than its usual number, and is accordingly charged positively. Indeed, in every instance in which dissociation occurs, *the ions are charged,* since one of the dissociated parts carries with it more than its usual number of electrons, and the other is left with a number correpondingly less. If it is a metallic salt or an acid which is thus dissociated by solution in water, it is invariably the acid part, or the *acid radical*, as it is sometimes called, that carries with it the excess of electrons and is charged *negatively*, while it is the *metallic part* in the case of salts, or *hydrogen* in the case of acids, that is charged *positively*.

389. Relation Between the Ionic Charges and the Chemical Bonds. Whenever one and only one chemical bond is thus broken by dissociation, as in the example of the NaCl mentioned above, one single electron in excess is taken by the acid radical, in this case the Cl, and a deficit of one electron is produced upon the positive ion, in this case the Na. If two chemical bonds are broken in this way, the acid radical carries with it an excess of two electrons, while the metallic or the hydrogen ion is left with a corresponding deficit of two electrons, and so on. The number of electrons which the negative ion carries is thus always the same as the number of the chemical bonds which have been broken by the dissociation, while the number of electrons which the positive ion is lacking is likewise the same as the number of these broken bonds. In other words, *when a substance dissociates, the negative ion carries with it an excess of one electron for each chemical bond that has been broken, while the positive ion is left with a deficit likewise of one electron for each bond thus broken.* The number of electrons which the negative ion carries in excess, and the corresponding number which the positive ion is lacking, are thus seen to have the most intimate and direct relation with the number of chemical bonds which, when dissociation occurs, would otherwise be left free.

390. The Passivity of the Ions. Inasmuch as ordinary chemical reactions take place when the atoms or radicals have their normal numbers of electrons, it might well be expected that these reactions would be different, or might possibly fail altogether, when the atoms or radicals thus become electrified and so form ions. Indeed, it is found that while the uncharged atoms and radicals are exceedingly active, so much so that ordinarily they cannot exist by themselves, they cease altogether to have these active chemical properties as soon as they acquire or lose one or more electrons; and therefore in their charged condition as ions, they float about in the water separated from each other.

We may form a conception of the ionic condition which then exists by imagining that the charges thus resident upon the ions possibly protect the bonds, one electron for each bond, and so prevent the activity which would otherwise take place. Certain it is that the presence or the absence upon an atom of one additional electron for each bond renders the atom passive and inert, and takes from it every trace of the chemical properties which that atom or radical would otherwise possess. Moreover, the ion will remain indefinitely in this inactive condition until it is discharged, that is, until electrons flow away from it or upon it and it once more has its normal number, after which, as the bonds are then no longer protected, it instantly regains its usual chemical properties.

391. Dissociation Complete Only in Dilute Solutions. It is true that in the solution the positive charge upon one ion and the negative charge upon another may, by their mutual attractions, bring the two together and for a moment form a molecule of the original substance. The same power of dissociation which effected the separation in the first place will, however, operate again, and the two will once more become ions. There are, therefore, always present in every solution a number of ions which become momentarily united, thus forming a neutral molecule, but which will again dissociate a moment later. From this follows the conclusion previously reached (p. 455) that the more dilute the solution the more complete the dissociation.

Just as NaCl dissociates into Na and Cl ions, the Cl being charged negatively and the Na positively, so hydrochloric acid, HCl dissociates into H and Cl ions, the Cl again having an excess, and the H a deficit, of one electron. In a similar way sulphuric acid, H_2SO_4, dissociates into two H ions and one SO_4 ion, the latter having two bonds, and therefore having upon it an excess of two electrons. The two H ions are accordingly charged positively, each having yielded an electron to the SO_4 radical.

Any solution containing a dissociated substance, that is, any solution containing ions, is called an electrolyte.

392. Local Action—Amalgamation. If a piece of ordinary commercial zinc is placed in dilute sulphuric acid, a violent chemical reaction takes place and hydrogen gas is evolved in large quantities. As a result of this action the zinc rapidly wastes away and a black residue consisting largely of carbon, which is the chief impurity in zinc, is all that remains. Indeed, the ordinary method of making hydrogen gas is to pour dilute acid upon pieces of zinc and to collect the hydrogen that escapes. When, therefore, a piece of ordinary zinc is used as one of the plates of a battery, this chemical reaction causes a rapid wasting away of the zinc, with the consequent evolution of hydrogen. This evolution of hydrogen at the surface of the zinc plate of a battery is called *local-action*. Since perfectly pure zinc, is not affected by dilute acid, in this way a plate of zinc wholly free from all impurities may be placed in this liquid without bringing about the evolution of hydrogen, and without decreasing appreciably in amount.

Local action in a battery, therefore, is an effect which in some way depends upon the presence of the impurities in the zinc, and this will be considered in detail later. These impurities may be covered up, and thus local action prevented, by pouring mercury over the zinc, thereby leaving the latter with a thin film of mercury adhering to it. The zinc, however, is soluble in the mercury, and hence zinc particles float through this covering and come in contact with the acid, so that a piece of zinc thus coated with mercury behaves in dilute acid precisely like a piece of chemically pure zinc. This process of overspreading the zinc with a film of mercury is called *amalgamating* the zinc. The amalgamation of the zinc, therefore, prevents local action by covering up the impurities, while it in no way detracts from the role which the zinc plays in the action of the battery.

393. The Formation of Zinc Ions in a Battery. When a plate of zinc is placed in water or in dilute acid, it tends to dissolve, and, indeed, it does dissolve to a certain extent. That is, a particle of zinc, which originally formed a portion of the zinc plate, separates from the plate and floats away by itself. Moreover, in the separation which is thus effected between the zinc particle and the original plate, it is found that the two chemical bonds which zinc possesses have been ruptured, and that two electrons from the zinc particle have been retained by the plate. In other words, the plate, having two electrons in excess, is thereby charged negatively and is at low potential, while the zinc particle, having a deficit of two electrons, is thus left charged positively and is at high potential. The zinc particle is therefore in every sense a zinc ion, inasmuch as it is a metallic particle which has lost one electron for each of its two broken bonds. A zinc ion, as well as a copper ion, which likewise has two bonds, thus carries a positive charge twice as great as that upon an H or a Na ion, since each of the latter has only a single bond.

394. Solution Tension is Analogous to Osmotic Pressure. If a plate of chemically pure zinc or a piece of amalgamated commercial zinc is placed in an electrolyte, the zinc dissolves and sends positively charged Zn ions into the solution. This makes the number of positively charged bonds in the solution in excess of the number of negatively charged bonds, and hence an electrostatic repulsion between these positively charged ions takes place. So enormously great is this electrostatic repulsion that the Zn ions already in solution repel those about to leave the zinc plate, and the zinc ceases to dissolve after a relatively small number of ions have been formed.

The zinc ions already in the electrolyte therefore exert a pressure, or tension, upon the zinc plate which prevents other particles of zinc from passing into solution. This electrostatic pressure of the postively charged ions against the zinc plate. thereby preventing the zinc from dissolving further, is analogous to the osmotic pressure (p. 230) of the particles of a dissolved substance against the separating membrane, tending to prevent a further passage of the water. Moreover, just as a definite osmotic pressure will prevent a further flow of water through the membrane, so a definite electrostatic pressure of the zinc ions already in solution will prevent more zinc ions from leaving the zinc plate and passing into the liquid. This electrostatic counter pressure, which thus opposes the tendency of zinc to dissolve, reaches the enormous figure of 10^{19} atmospheres.

Accordingly, *the tendency of a metal to dissolve, that is, the solution tension of that metal, is the highest possible electrostatic counter pressure under which the metal is able to send ions into the solution.* The solution tension of zinc, therefore, is inconceivably great; and when the zinc ceases to dissolve, as it does very quickly when placed in an electrolyte, it indicates that the zinc ions which are already dissolved, exert the above enormous pressure, analogous to an osmotic pressure, upon the surface of the zinc plate.

If now, a piece of copper is placed in the electrolyte near the zinc, it will be found that the copper has a solution tension which is only 10^{-20} atmospheres, hence the number of ions which it can send into the solution is infinitely small as compared with those which come from the zinc, and therefore these may be neglected altogether. The copper, however, is a conductor; and so, when a positively charged ion touches it, an electron flows from it off upon each positively charged bond of the ion. This discharges the ion, or makes it electrically neutral, so that it becomes a free chemical atom with free, active bonds. Since a free atom cannot exist by itself, this atom then unites with another similar atom to form a molecule, which is deposited upon the copper plate; and thus, in passing out of the solution, it removes a positive ion and makes it possible for another particle of zinc to leave the zinc plate and become an ion.

395. The Action of the Simple Battery Cell. It is now possible to study somewhat in detail the processes which take place within the battery.

If the fluid of the battery is dilute sulphuric acid, the H_2SO_4 dissociates at once, each molecule yielding one negatively charged SO_4 ion and two positively charged H ions. The SO_4 ion, having two chemical bonds, carries an excess of two electrons, one for each bond, while each of the two hydrogen ions is lacking an electron, as H has a single chemical bond only. The electrolyte, however, is absolutely neutral, since the number of the positively charged H bonds equals the number of the negatively charged SO_4 bonds. Into this electrolyte we bring a plate of copper (Cu) and a plate of Zinc (Zn), Fig. 250, the latter being the more massive, since this is the plate that wastes away and will ultimately disappear through the action of the battery.

The processes which now takes place in the battery cell, whereby the zinc becomes at low potential and the copper at high potential may be described as follows: The zinc sends Zn ions into the solution until the electrostatic pressure of these ions against the Zn plate is equal to the solution tension of the zinc. At the separation from the zinc plate, each Zn ion yields two electrons to the plate, thereby charging the latter negatively and bringing it to low potential. The low potential of the zinc plate is thus occasioned by the accumulation upon it of the electrons which it retains when the zinc ions go into solution.

The presence in the electrolyte of the positively charged Zn ions produces the two-fold effect of attracting the negatively charged SO_4 ions and repelling the positively charged H ions. There is, therefore, throughout the liquid a steady though gradual movement of the SO_4 ions towards the region of the zinc plate, and also a movement of the H ions in the opposite direction, that is, towards the region of the copper plate.

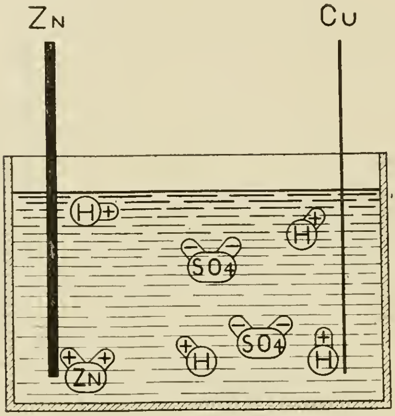

Fig. 250. The action of the simple battery cell.

A Zn ion will ultimately attract to itself an SO_4 ion, thus forming a neutral molecule of zinc sulphate, $ZnSO_4$. As long as the solution of this salt remains dilute, this newly-formed molecule will at once dissociate, forming again a Zn and an SO_4 ion; but as these will now remain in the immediate neighborhood of each other, they produce together a resultant electrostatic field of zero intensity. As the solution of $ZnSO_4$ becomes more concentrated, this salt ceases to dissociate and even forms crystals, which so impede the action of the battery that the old electrolyte must be removed and replaced by new.

It is, however, the positively charged H ions in the electrolyte that are effective in bringing about the high potential of the copper plate. As was shown above, the positively charged Zn ions not only attract to themselves the negatively charged SO_4 ions, but also repel the positively charged H ions. These then repel other H ions, and these latter still others, and so on, until a chain of these ions is formed extending through the electrolyte to the copper plate, each ion repelling the one before it, and being in turn repelled by the one behind it. Finally, for each Zn ion that goes into solution at the zinc plate, two H ions near the copper plate move slightly forward and touch this plate, which, as we have seen, is a conductor. Two electrons then flow from the copper plate off upon these two H ions, thus making the latter neutral chemical hydrogen atoms, and leaving the copper plate charged positively, and therefore at high potential.

It is, however, impossible for free hydrogen atoms to exist by themselves; and as two atoms are necessary to form a molecule, these two hydrogen atoms unite with each other by their free chemical bonds, thereby forming a molecule of hydrogen which is deposited upon the copper plate. This removal from the electrolyte of two positively charged hydrogen ions decreases the electrostatic repulsion against the zinc plate, so that another Zn ion, with its two positively charged bonds leaves the zinc and passes into the solution. All of the above reactions are then repeated in the same order.

This process, which thus seems continuous, is, however, limited by the potential which the copper plate soon attains. The potential of this plate continues to rise as electrons flow from it to the H ions, until it becomes equal to the potential of the ions themselves. All action then ceases; and the two plates, the copper at high potential and the zinc at low potential, remain in the electrolyte in this condition of equilibrium.

It is in this condition, that is, when the copper and the zinc are not connected by a wire, that the electromotive force of a battery is measured, for it is only when everything has thus come to a state of equilibrium that the true effect of these processes which go on in the battery may be determined. The electromotive force of a battery, therefore, might perhaps more properly be said to be due to a difference in the solution tension of zinc and copper, rather than to what is ordinarily known as chemical action. A battery could not be operated with two metals which had the same solution tension in a single electrolyte.

396. The Flow of the Electrons. If the zinc is now connected by a wire with the copper, electrons will flow through the wire from the low potential zinc to the high potential copper, thus raising the potential of the zinc and lowering the potential of the copper. The fact that the copper is now at lower potential makes it possible for still other hydrogen ions to approach and touch this plate, and to receive electrons from it. This further flow of electrons from the copper plate both raises the potential of this plate and also discharges the hydrogen ions upon which the electrons flow, so that these ions become neutral chemical atoms, which, when they unite with other similar atoms and form molecules of hydrogen, are deposited upon the copper plate.

This removal of H ions from the electrolyte decreases the electrostatic repulsion against the zinc plate, so that more zinc can now dissolve and pass as ions into the solution, thereby once more lowering the potential of the zinc plate. Again a flow of electrons will take place from the zinc through the wire to the copper, again electrons will flow from the latter plate off upon H ions which approach and touch it, again hydrogen molecules will be deposited upon the copper plate, and so the operation of the battery becomes continuous. So rapidly, indeed, do these processes take place that the $ZnSO_4$ which has not yet been dissociated may be seen trailing off from the lower end of the zinc plate; while hydrogen molecules accumulate in such numbers upon the copper plate that they unite to form bubbles of hydrogen gas, which rise through the liquid in a continuous stream.

The metallic connection between the zinc and the copper plate may be made beneath the surface of the liquid quite as well as above it.

397. The Flow of the Current. This flow of electrons through the wire, as described above, is what is understood to be the flow of the current. In this connection, however, attention may well be called once more to the unfortunate fact that in scientific as well as popular parlance, the current is said to flow from the copper through the wire to the zinc. As has been pointed out in a previous chapter (p. 351), this is a legacy that has come down to us from the time when it was thought that the positive charge moved or flowed; and although we now know that it is the negatively charged electrons that flow through the wire from the zinc to the copper, we are still, and probably shall always be, constrained to speak of the current as flowing from the high potential copper to the low potential zinc.

It has likewise been pointed out that the *exterior effect* of a negative charge moving in the negative direction is the same as that of a positive charge moving in the positive direction. We may therefore avoid all ambiguity and all confusion by discriminating carefully between *the direction of flow of the electrons*, and *the direction of flow of the current*; and when the *current* is mentioned as flowing in a particular direction, we may always understand that what is actually taking place is that the *electrons* are flowing in the opposite direction.

398. The Explanation of Local Action. We are now in a position to understand the true nature of local action in a battery (p. 458), and to explain the rapid evolution of hydrogen gas when ordinary zinc is treated with dilute sulphuric acid. The solution tension of the particles of the carbon, which forms the chief impurity in commercial zinc, is even less than that of copper, and at the same time carbon is an excellent conductor of electricity. When, therefore, the impure zinc, with its particles of carbon adhering to it, is placed in an electrolyte, the zinc tends to dissolve just as if it were chemically pure, and the Zn ions which thus go into solution, tend to repel H ions and force them out of the solution.

By touching the carbon particles, the H ions receive electrons and so become discharged, Fig. 251, just as the H ions receive electrons from the copper plate in the action of the battery already described. As soon as these ions are discharged, they become chemical hydrogen atoms, and unite with other similar atoms to form hydrogen molecules, which in turn gather into bubbles and rise in great profusion from the surface of the carbon particles. In other words, every particle of carbon upon the face of the zinc acts precisely like the copper plate of a battery; and an evolution of hydrogen takes place upon this particle for the same reason that hydrogen is evolved upon the copper plate.

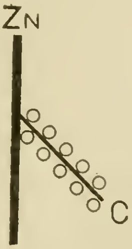

FIG. 251. Local action.

There are, therefore, myriads of little batteries all up and down the face of the zinc, each producing a tiny current flowing from the carbon through the point of contact to the zinc. These little local currents not only produce an evolution of hydrogen gas from the surface of the carbon particles, but also cause the zinc to waste away, as this always does when the current is flowing. Therefore, if local action were allowed to continue, the zinc would soon waste entirely away, being consumed thus in maintaining these little parasitic currents which exist between the impurity and the zinc wherever a particle of impurity is found. Accordingly, if these particles of impurities are covered up by a coating of mercury, in which the zinc dissolves readily, the zinc may still be brought in contact with the electrolyte, while the impurities, which would otherwise form the high potential plates of innumerable little batteries, are surrounded by a covering which effectually prevents them from coming in contact with the battery fluid. Thus it is that the amalgamation of the zinc greatly increases the efficiency of the battery.

399. Polarization of the Battery. We have seen that the action of the battery is inevitably such as to deposit positively charged ions upon the copper plate. If these are hydrogen ions, they will lose their charges of electricity and gather into bubbles, thus covering the copper plate and producing the effect known as the *polarization* of the battery. If the battery is allowed to polarize, that is, if these bubbles of hydrogen are allowed to collect upon the copper plate, the difference of potential between the copper and the zinc is greatly diminished, and the flow of current may even cease altogether. Polarization, therefore, is a phenomenon most detrimental to the successful action of the battery, and every effort is made to obviate this objectionable feature.

If the bubbles of hydrogen are allowed to gather upon the copper plate, this plate is then partly covered by a layer of gas, so that its effective portion, that is, the portion in contact with the electrolyte, is greatly diminished. As we shall learn later, diminishing the size of the plates of a battery increases the resistance which the flow of the electricity has to encounter; and therefore, the flow of current decreases in a battery which is allowed to polarize, due in part to the increased internal resistance.

A far more serious objection, however, lies in the fact that the resulting difference of potential between the copper and the zinc is very much less in a battery where polarization is allowed to take place. Indeed, the solution tension of hydrogen is almost identically the same as that of zinc, so that the hydrogen bubbles which coat the copper plate tend to dissolve back again in the electrolyte and become ions once more.

As each hydrogen particle thus dissolves and becomes a positively charged H ion, it parts with an electron which is retained by the copper plate, whereby the potential of this plate is lowered; and as this process continues, the potential of the copper plate falls lower and lower, until it is but little, if any, above the potential of the zinc plate. Accordingly, the effective difference of potential between the zinc plate and the copper plate is very greatly decreased, and the flow of current in such a battery soon becomes very feeble, or may, indeed, cease altogether, due to the fact that the difference of potential between the plates has been destroyed.

The decrease in potential of the copper plate, due to this tendency of the hydrogen bubbles again to dissolve, is called *the counter electromotive force of polarization*. That is, a copper plate coated with hydrogen bubbles, and therefore behaving in a battery almost exactly like a zinc plate, tends to develop an electromotive force in a direction opposite to that which is produced between the zinc and the copper when the battery operates without polarization. The resultant electromotive force in a battery which is polarized is therefore only the difference between the direct electromotive force of the battery and the counter electromotive force of polarization, and this difference usually becomes very small indeed.

400. The Method of Avoiding Polarization—The Two-Fluid Battery. In order to avoid polarization, it is necessary to surround the copper plate by an electrolyte that does not contain hydrogen ions, and this can be accomplished only by the two-fluid battery. While there are many forms of two-fluid batteries, all being devised for the sole purpose of avoiding polarization, only one will be described in detail, but this may be taken as a type of all.

401. The Daniell Battery. The Daniell battery consists of a plate of zinc immersed in dilute sulphuric acid, H_2SO_4, and of a copper plate surrounded by a solution of copper sulphate, $CuSO_4$, these two liquids being separated by a porous partition, Fig. 252. This porous partition, which consists of unglazed earthenware, is close enough to prevent the two liquids from mingling, while it is still sufficiently open to permit ions to wander through its pores. The H_2SO_4 dissociates upon being dissolved in water into two H ions, and one SO_4 ion. In the process of dissociation, the SO_4 ion, which has two bonds, has deprived each of the H ions of an electron, and so is charged negatively, while the two H ions are thus left charged positively, inasmuch as each has a single bond and has lost one electron. In a similar manner, the $CuSO_4$ molecule dissociates into one negatively charged SO_4 ion and one positively charged Cu ion, the latter having two chemical bonds, and therefore being left with a deficit of two electrons. A zinc plate is now suspended in the H_2SO_4 electrolyte, and a copper plate is immersed in the solution of $CuSO_4$.

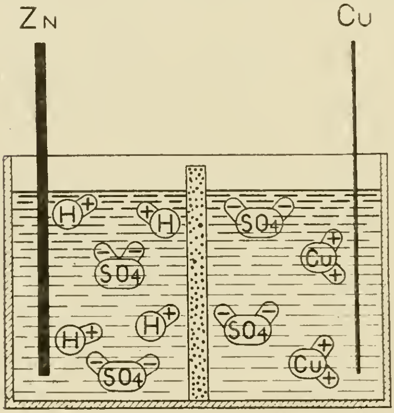

FIG. 252. The principle of the Daniell battery.

As before, the great solution tension of the zinc causes positively charged Zn ions to separate from the zinc plate and go into solution. The electrostatic repulsion between these Zn ions and the H ions which are already in the solution now takes place, each Zn ion repelling two H ions, and these in turn repelling two others, and so on, until the porous partition separating the two fluids is reached. These H ions now pass through the porous partition and enter the $CuSO_4$ electrolyte. Here the electrostatic repulsion continues, since this side now contains an excess of positively charged ions. Therefore the two H ions repel a Cu ion, this in turn repels another Cu ion, this another, and so on, until finally a Cu ion touches the copper plate. This Cu ion now receives two electrons from the copper plate and becomes itself a free, chemical, copper atom with its two free, active bonds. Two such atoms of copper now join together by their free bonds and form a molecule of copper, which is deposited upon the copper plate.

By the action of this two-fluid battery, therefore, *a molecule of copper instead of a molecule of hydrogen is deposited upon the copper plate;* but since this was already a copper plate, the addition of more copper can in no way change its solution tension, and therefore the full difference of potential between the zinc and the copper is maintained *without the least trace of a counter electromotive force due to polarization.* The only condition for the successful operation of this battery is to keep the copper plate continuously immersed in a saturated solution of $CuSO_4$, so that the H ions can never approach it. To this end the copper plate is usually surrounded by crystals of $CuSO_4$, which keep the liquid saturated with this salt. As soon, therefore, as an ion of copper is repelled out of the solution and deposited as metallic copper upon the copper plate, another particle of the $CuSO_4$ dissolves; thus the copper plate is continually surrounded by positively charged Cu ions which effectually prevent the H ions from approaching it and depositing themselves upon it. Thus, by the employment of a two-fluid battery, polarization is avoided.

402. The Gravity Battery. Another form of the Daniell cell, known as the gravity battery, is more convenient and is of more frequent occurrence. In the gravity battery no porous partition divides the two electrolytes, Fig. 253, but this separation is accomplished by the greater specific gravity of the $CuSO_4$ solution, which causes it to remain in the lower part of the battery jar.

Accordingly, a copper plate is placed in the bottom of the jar which is then partly filled with a saturated solution of $CuSO_4$, and crystals of the same salt are placed in large quantities around this copper plate. Upon this solution of $CuSO_4$ is carefully poured the lighter solution of dilute sulphuric acid, or, what is quite as good, a dilute solution of $ZnSO_4$, and in this, in the upper part of the jar, is immersed the plate of zinc.

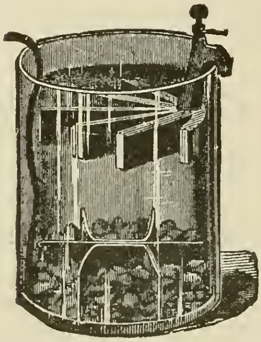

FIG. 253. The gravity battery.

The mingling of these two solutions is prevented by keeping a closed circuit, through which the current continually flows. This gives *direction* to the motion of the ions, so that the Cu ions are confined to the lower part of the jar in contact with the copper plate, as shown by the blue color of this solution, while the H ions, as well as the Zn ions, remain in the colorless electrolyte above. This separation of the two fluids by gravity is the easier to accomplish from the fact that a solution of $CuSO_4$ mingles by diffusion very slowly indeed with the clear overlying liquid. Even this tendency toward diffusion would ultimately effect a mingling of the two liquids, were it not for the fact that the Cu ions are repelled downwards toward the copper plate by the H and the Zn ions of the upper electrolyte. The great specific gravity of a $CuSO_4$ solution and the slow rate of diffusion which this liquid possesses are the agents which are effectual in keeping apart the two liquids of the gravity battery.

403. The Grove Battery. Another form of the two-fluid battery which has proved most serviceable in the past, although it is used but little at the present time, is the Grove battery. In this cell also a porous partition separates the two electrolytes, the zinc in a solution of H_2SO_4 being on one side, and a strip of platinum immersed in strong nitric acid being on the other. The platinum has a solution tension which is even less than that of copper, and hence the electromotive force of this cell is slightly greater than that of the Daniell battery.

The action of the Grove cell is similar to that of the Daniell battery, in that the zinc dissolves in the H_2SO_4 electrolyte, an ion of Zn repelling two H ions, and these two others, and so on, until two H ions pass through the porous partition. On the other side of this partition these two H ions repel two other H ions, these two repel still two others, and so on, until finally two H ions touch the platinum plate. Two electrons then flow from the platinum plate upon these ions, thereby raising the potential of the plate, and changing the ions into free H atoms. These free H atoms, with their free, active, chemical bonds, here encounter the strong nitric acid, which is a powerful oxidizing agent, and this oxidizes the two H atoms, an atom of oxygen from the nitric acid uniting with these two H atoms to form a molecule of water H_2O. Hydrogen bubbles, therefore, cannot collect upon the platinum plate, because they are oxidized as soon as formed, and hence polarization is prevented.

Although this battery is most efficient, having a high electromotive force and a low internal resistance, it has been but little used owing partly to the great expense of the platinum plates, and partly to the corrosive fumes which are produced by the nitric acid when this is reduced in the process of oxidizing the hydrogen. The first objection is practically removed in the Bunsen battery, where the platinum is replaced by a plate of carbon; but the corrosive acid fumes are so deleterious that even this excellent battery has found but little favor.

404. The Bichromate Battery. A battery more successful for general use is the bichromate battery, which consists, in its essentials, of a plate of zinc in an H_2SO_4 electrolyte, and a plate of carbon also in a solution of H_2SO_4, to which, however, a quantity of potassium bichromate has been added. The bichromate solution is a fairly good oxidizing agent, and by this means the hydrogen is oxidized, as in the Grove and the Bunsen batteries, thus avoiding polarization. Although the fumes are thus eliminated, the bichromate solution soon becomes exhausted, so that the action of this battery is somewhat limited.

405. The Leclanche Battery. The Leclanché, or door-bell battery, is a single-fluid battery, containing a zinc and a carbon plate in a solution of ammonium chloride. In some forms, the ineffectual attempt to prevent polarization is made by surrounding the carbon plate with manganese dioxide, which to a small extent oxidizes the hydrogen. Even under the most favorable conditions, however, the battery polarizes rapidly, so that it must be used exclusively for open-circuit work, that is, for intermittent currents or currents of very short duration, such as those which are used in ringing a vibrating bell.

It may be said, in general, that primary batteries of all forms are rapidly going out of use, owing to the prevalence of the more convenient storage battery, and the universal introduction of the dynamo for purposes of power and light.

406. The Practical Unit of Electromotive Force—The Volt. We have seen that, in view of the solution tensions of zinc and copper, a perfectly definite difference of potential exists between the two plates of a battery. Although batteries differ greatly in their construction and materials, the electromotive forces of all good batteries are just about the same, and do not differ greatly from that of the Daniell cell. With the introduction of batteries, came also the demand for a unit in terms of which to measure and compare electromotive forces, and it was seen that the most convenient unit would be one which was as near as possible the electromotive force of the Daniell battery. Inasmuch as the electrostatic unit of potential (p. 356) is about three hundred times greater than the electromotive force of the Daniell cell, a unit equal to $\frac{1}{300}$ of this electrostatic unit was adopted as the working unit of electromotive force, and this, in honor of the Italian discoverer Volta, was called the volt. The volt therefore is the practical or working unit of electromotive force in current electricity, and may be provisionally defined as a unit equal to $\frac{1}{300}$ of the electrostatic unit of difference of potential.

For reasons which will be pointed out later, still another system of electrical units, known as the electromagnetic, is in common use, and it is to the units of this system that all measurements which have to deal with dynamic or current electricity are referred. Besides the volt, then, there is also an electromagnetic unit of electromotive force, which will be defined in terms of C. G. S. units at another time. This unit, however, is exceedingly small; and because of its extreme minuteness a multiple by 100,000,000 was taken arbitrarily in order to obtain a working unit which should be a multiple by 10 of the electromagnetic unit, and still represent as nearly as possible the electromotive force of the Daniell cell. The volt is therefore defined as 10^8, or 100,000,000, C. G. S. electromagnetic units of electromotive force.

407. Batteries Connected in Series. Since a flow of electrons is produced by a difference of potential, the flow may be increased, that is, the current may be made stronger, other things being equal, by increasing the difference of potential between the terminals of the battery. It must be remembered, however, that the differences of potential which occur in batteries are relatively very small indeed, and that it is most desirable, if possible, to increase this difference.

At first, it might seem that making a battery with very large plates would produce a greater difference of potential, but upon testing this we find that the electromotive force of a battery with very large plates is no greater than that of a battery with very small plates. A second reflection will show, indeed, that this must be so, for we have seen that the electromotive force of a battery depends upon the solution tension of the two metals in the fluids of the battery, and this is just as great for a battery with very small plates as for one with very large plates.

Accordingly, we see that *the electromotive force of a battery depends altogether upon the nature of the plates and the kind of fluids in the battery* into which these are placed, and does not depend at all upon the size of the plates, or their distance apart, or the quantity of the liquid or liquids present. It is true that it is desirable to have batteries with large plates, but this is for the purpose of decreasing the resistance to the flow of electricity, and also for providing plates that will not soon become exhausted by the action of the battery; but the great fact is constantly to be borne in mind that the electromotive force of a battery is absolutely independent of the shape, or size, or distance apart, of the plates, and *depends only upon the material of the plates and the nature of the liquids in which they are placed.*

It is possible, however, by using several batteries, so to join them that the difference of potential between the first and the last may be increased. If the copper of the first cell is joined to the zinc of the second, the copper of the second cell to the zinc of the third, and so on, the cells are said to be joined *in series*. *When n similar cells are joined in series, the electromotive force of the combination is n times the electromotive force of a single cell.* That this is so may be seen by examining the potentials in each cell throughout the series.

If we imagine the zinc of the first cell to be connected with the earth, Fig. 254, its potential is zero. By the action of the cell, the copper is always at a higher potential than the zinc, and we will assume this difference to be 1 volt. Accordingly, if the potential of the first zinc is zero, the potential of the copper in the first cell is 1 volt. This copper, however, is connected by a wire with the zinc of the second cell, and, therefore, the zinc of the second cell is at 1 volt. But the copper in a cell is always 1 volt higher in potential than the zinc of that cell, therefore the potential of the copper in the second cell is 2 volts. This copper is now connected by a wire to the zinc of the third cell, hence this zinc is at a potential of 2 volts, and the copper is 1 volt higher, or at 3 volts. In the same way, the copper of the third cell is connected by a conductor to the zinc of the fourth, hence this is at a potential of 3 volts, while the copper of the fourth cell is at a potential of 4 volts. The difference of potential between the zinc of the first cell and the copper of the last cell is, in this instance, 4 volts, or 4 times the difference of

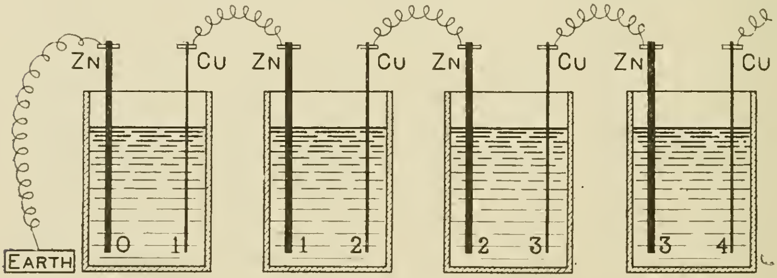

FIG. 254. Batteries connected in series.

potential of a single cell. *By joining cells in series, therefore, a difference of potential is obtained equal to that of a single cell multiplied by the number of cells so joined.*

408. Batteries Connected in Multiple, or Parallel. There is, however, another method of joining cells, called connecting them *in multiple, or in parallel.* When cells are joined in multiple, all the copper plates are connected together and all the zinc plates are likewise connected together, Fig. 255. A flow of current is then obtained by connecting any point of the wire which unites the copper plates with any part of the wire which joins together the zinc plates.

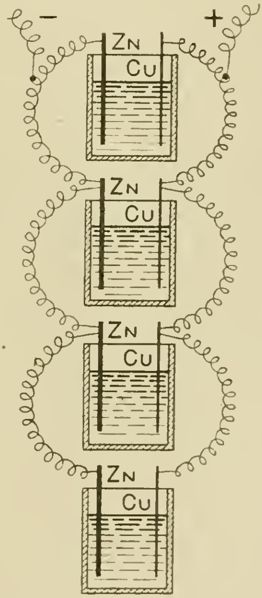

Fig. 255. Batteries connected in multiple, or parallel.

Any number of cells connected in multiple, however, produce only the electromotive force of a single cell. This becomes at once obvious by considering that if the zincs are connected by a wire they must all assume a common potential, and if the coppers are connected by a wire they must likewise be at one potential. Inasmuch, however, as the copper in any cell is 1 volt higher than the zinc of that cell, the potential of the whole number of the coppers, which are all connected together, can be only 1 volt higher than the potential of the entire number of zincs, which are also connected together. It is seen, therefore, that *the electromotive force of any number of battery cells connected in multiple is just the same as the electromotive force of a single cell*, that is, the electromotive force remains unchanged by connecting cells in multiple.

This may be seen from another point of view by remembering that in a battery joined in multiple the zinc plates, which are all connected together, act as one great zinc plate, and the copper plates, which are all connected together, act as one great copper plate; and since the electromotive force is independent of the size of the plates, it follows that, whether the battery consists of one cell or of many cells connected in multiple, the difference of potential remains just the same.

A striking illustration of the absolute independence of the electromotive force of a battery upon the size of the plates is obtained by making a cell of a copper gun-cap and a slender strip of zinc, moistened with a single drop of water. If the plates and the liquid are of the same nature as those in a battery with much larger plates, the electromotive forces of the two will be found to be the same.

CHAPTER XXXIII.

ELECTROLYSIS.

409. Electrolysis. If two platinum plates are placed a little distance apart in an electrolyte and connected with the poles of a battery, the action of the battery will maintain these plates at a constant difference of potential, so that the plate connected with the copper will be charged positively and be at high potential, and that connected with the zinc will be charged negatively and therefore be at low potential. These charged plates now exert electrostatic attractions and repulsions upon the ions of the electrolyte. The high potential plate will repel positively charged ions and attract negatively charged ions, and the low potential plate will repel negatively charged ions and attract positively charged ions.

Thus, from the fact that two plates, one at high potential and the other at low potential, are placed in the electrolyte, *the motion of the ions is directed*, the positively charged ions moving toward the negatively charged plate and the negatively charged ions moving toward the positively charged plate. In this way direction is given to the motion of the ions which thus form a double procession, those charged positively all moving one way and those charged negatively moving the other. When these charged ions finally touch the plates towards which they are moving, they become discharged, one electron for each bond flowing from the negatively charged ion to the high potential plate, and likewise one electron for each bond flowing from the low potential plate to the positively charged ion. The ions then become free chemical atoms; and since a single chemical atom seldom remains by itself, one such atom joins with another to form a molecule. Thus a molecule of one part of the dissociated substance is deposited upon one plate, and a molecule of the other part of the dissociated substance is deposited upon the other plate. In this way, by the action of a battery which maintains a constant difference of potential between these plates, *the constituent parts of the electrolyte are taken out of solution and deposited upon the two plates.*

This process, which is called *electrolysis, consists* therefore, *in discharging the ions of an electrolyte and depositing them as chemical substances upon plates connected with a battery.* The two plates upon which these chemical substances are deposited are known as the *electrodes*, the one at high potential being called the *anode*, and the one at low potential the *cathode*. The tank itself in which this decomposition takes place is called an electrolytic, or decomposing cell.

410. The Electrolysis of Dilute Sulphuric Acid—The Decomposition of Water. In order to follow this process of electrolysis in a specific case, let us assume that the electrolyte is dilute sulphuric acid, H_2SO_4. We then have present positively charged H ions, each having one chemical bond, and negatively charged SO_4 ions each with two chemical bonds. The positively charged H ions now move toward the cathode, Fig. 256, and the negatively charged SO_4 ions move towards the anode. These ions are discharged as soon

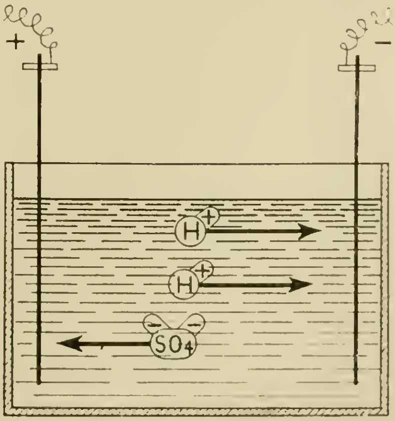

Fig. 256. The movement of the ions in electrolysis.

as they touch their respective electrodes, and so become free chemical particles. Two hydrogen ions are thus deposited upon the cathode for every single SO_4 ion that is deposited upon the anode. An atom of H unites with another atom of H to form a molecule, and molecules collect to form a bubble which appears as such upon the cathode.

The SO_4 at the anode, however, is a very unstable chemical compound. As long as it remains charged, it is perfectly inactive, but as soon as it is discharged by touching the anode, it breaks up into O and SO_3, according to the reaction

$$SO_4 = SO_3 + O.$$

The sulphur trioxide, SO_3, which is chemically very active, now unites with a molecule of water, H_2O, as follows:

$$SO_3 + H_2O = H_2SO_4.$$

The molecule of H_2SO_4 thus formed is at once dissociated into two H ions and one SO_4 ion, and therefore the number of these ions in the electrolyte is now just as great as at first. The atom of O, which is thus set free by the breaking up of the SO_4 radical, unites with another atom of O to form a molecule of oxygen, and the molecules form a bubble which appears upon the anode.

Accordingly, while hydrogen gas bubbles up at the cathode, oxygen gas is given off from the anode. Moreover, it is to be noticed that for every two molecules of hydrogen which thus appear at the cathode, only one molecule of oxygen is formed at the anode, or in other words, the volume of hydrogen which is thus set free is twice as great as the volume of oxygen. In the process of the electrolysis of dilute H_2SO_4, therefore, while it is true that the two parts of this substance appear at the electrodes, the *secondary* reactions which take place at the anode cause a molecule of water to break up and again to form a molecule of H_2SO_4. It is, therefore, ultimately the H_2O which is decomposed by electrolysis, while the amount of H_2SO_4 remains perfectly constant in the solution. Thus by electrolytic processes it is seen that water consists of two volumes of hydrogen combined with one volume of oxygen.

411. Counter Electromotive Force Due to Electrolysis. Investigation shows that the presence of the oxygen and the hydrogen bubbles which gather upon the electrodes produce a counter electromotive force somewhat greater than that of the Daniell battery. Since this counter electromotive force opposes the action of the direct electromotive force, it is evident that a battery having a higher electromotive force than the Daniell cell is necessary to decompose water. Accordingly, one Daniell battery will not electrolyze water, while two in series will do so readily.

412. The Electrolysis of Various Salts. If copper sulphate, $CuSO_4$, is subjected to electrolysis, the positively charged Cu ions will move towards the negative cathode and be deposited as metallic copper upon it, while the negatively charged SO_4 ions will move toward the positive anode and be for an instant deposited upon it. The secondary reaction at this electrode, however, as described above, will at once take place, according to the following reaction:

$$SO_4 = SO_3 + O$$

$$SO_3 + H_2O = H_2SO_4$$

$$H_2SO_4 = H,\ H,\ SO_4\ \text{(ions)};$$

so that while metallic copper is deposited upon the cathode, oxygen gas will bubble up at the anode.

If the substance to be electrolyzed is sodium chloride, NaCl, the Na will appear as metallic sodium at the cathode and the Cl will be deposited as gaseous chlorine at the anode. If, now, the electrolyte is colored red with litmus, the chlorine, which appears at the anode, and which is a strong bleaching agent, will destroy all trace of color, so that the solution around the anode becomes colorless. At the cathode, however, a secondary reaction takes place due to the fact that the free metallic sodium which is deposited upon this electrode cannot exist as such in the presence of water. The sodium, accordingly, attacks a molecule of water according to the following reaction:

$$Na + H_2O = NaOH + H.$$

The NaOH, which dissociated at once into Na and OH ions, is an alkali, and so changes the litmus solution from red to blue. Accordingly, if a solution of NaCl is reddened with litmus and subjected to electrolysis, chlorine gas will be given off at the anode and hydrogen gas at the cathode, while at the same time the color will disappear in the liquid around the anode, and will change from red to blue about the cathode.

If a solution of sodium sulphate, Na_2SO_4, is subjected to electrolysis, the Na_2 will be deposited upon the cathode and the SO_4 will appear at the anode. In this case secondary reactions will take place at both electrodes. At the anode, the SO_4 breaks up into SO_3 and O, the O being set free and the SO_3 at the same time uniting with a molecule of H_2O to form H_2SO_4. At the cathode the Na_2 unites with two molecules of H_2O, forming 2NaOH and setting free 2H. Accordingly, the solution around the anode becomes acid and will turn litmus red, while that around the cathode becomes alkaline and will turn litmus blue. Moreover, O appears at the anode and H at the cathode in the proportion of two volumes of H to one volume of O.

In all these examples of electrolysis it is to be noticed that in every instance the metallic and the H ions are positively charged, and therefore *the metals and hydrogen are always deposited upon the negative plate.* Moreover, the acid radicals in the electrolyte, such as SO_4 and Cl, are always negatively charged, and therefore *the acids and also oxygen are always liberated at the positive plate.*

A striking experiment in electrolysis, in which the metallic ions may be seen to deposit themselves upon the negative electrode in wonderfully beautiful and delicate groups, is performed when a current of electricity is sent through a solution of the sugar of lead. The positively charged lead ions move towards the cathode, and are deposited as metallic lead upon it. The lead, however, does not form an even coating upon this electrode, but is deposited in a series of crystals, which are seen to grow out from the cathode in delicate, fern-like branches, these branches having other branches, and so on, until the whole forms a beautiful structure which has been most appropriately called the lead tree.

413. The Flow of Current Through an Electrolyte. Throughout these illustrations of electrolysis, and, in truth, in every instance in which a directed motion is imparted to the ions of an electrolyte, this one all-important and most fundamental fact must constantly be borne in mind, that *the directed motion of the ions in an electrolyte is the current of electricity flowing through the electrolyte.* This is simply saying that *a current of electricity cannot flow through an electrolyte except as it is carried bodily upon the ions.* Accordingly, the negatively charged ions move towards the high potential electrode and give to it one election for each chemical bond, while the positively charged ions move towards the low potential electrode and receive from it likewise one electron for each bond. The electrons therefore flow from the negatively charged ions off upon the anode on one side of the electrolytic cell, and from the cathode off upon the positively charged ions on the other side of the cell. *Each ion as it moves through the electrolyte acts as a little carrier, and carries with it its electrical charge; and this directed movement of these charged ions, the positive ions, all moving one way, and the negative ions all moving the other, is the current of electricity flowing through the electrolyte.*

An aqueous solution of a substance cannot conduct electricity unless ions are present, pure water itself, which contains at most only a small number of ions due to its own dissociation, being practically a non-conductor. Indeed, water has been so completely freed from all ions of dissolved substances, and thereby rendered so poor a conductor, that electricity will experience greater resistance in passing through a column of water one millimeter long than in passing a thousand times around the earth through a copper wire of the same cross-section.

414. Provisional Definition of the Unit Current—The Ampere. In the work which has preceded, we have spoken of a flow of electricity, or the electric current, without, however, giving attention to the method by which this flow is measured, or to the meaning attached to the terms, *the strength, or the* *intensity, of the current.* These are expressions of great importance and of frequent occurrence in the study of electricity, and will be defined in terms of C. G. S. units at a later time. It will perhaps suffice to say here, however, that *the strength of the current is measured in terms of a unit called the ampere*, where the ampere may be *provisionally* defined as the current which in a unit of time will deposit a certain amount of a particular metal upon the cathode of an electrolytic cell. Copper is the metal ordinarily used, though on many accounts silver is preferable, as it gives results of greater accuracy. Accordingly, a current of 1 ampere is a current of such strength that it will deposit 0.0011180 grams of silver, or 0.0003281 grams of copper in one second.

In its proper place, another unit of current, known as the electromagnetic unit will be defined, and, indeed, it is from this unit that the ampere has been derived. The electromagnetic unit of current is relatively a very strong current; and in order to obtain a unit that should more nearly equal the currents ordinarily used, $\frac{1}{10}$ of this unit was arbitrarily taken as a practical or working unit, and was called the ampere. The ampere, therefore, is defined as $\frac{1}{10}$ of the electromagnetic unit of current.

415. The Unit of Quantity—The Coulomb. Having thus, at least provisionally, defined the ampere, which is the working unit of current, it is now possible to determine another electrical unit, namely, that of *quantity*. *The practical unit quantity of electricity in the electromagnetic system is called the coulomb, and is defined as the total quantity of electricity which passes through a conductor when a constant current equal to 1 ampere flows for 1 second.* The coulomb, therefore, is an ampere per second, or

$$\text{coulombs} = \text{amperes} \times \text{seconds}.$$

Thus, the coulomb is the unit quantity of electricity derived from the electromagnetic system, and so is not the same as the electrostatic unit of quantity previously defined (p. 339). Indeed, as it would require 3×10^9, or 3,000,000,000 electrostatic units of electricity to equal 1 coulomb, it is seen that the actual quantity of electricity developed by batteries is enormously greater than that to which the ordinary phenomena of electrostatics are due.

416. The Laws of Electrolysis. It is to the great English physicist, Faraday, that we owe our knowledge of the quantitative laws according to which electrolysis takes place. These laws, which are three in number, and which are still known as Faraday's laws, may be enumerated as follows:

1. The amount of substance deposited upon the electrodes by a current of given strength is proportional to the time of flow of the current.

2. The amount of substance deposited in a given time by the current is proportional to the strength of the current.

3. If the same current is sent through several electrolytic cells in series, the amount of substance deposited in each cell is proportional to the chemical equivalent of that substance.

417. Discussion of the First Law. As regards Faraday's first law, it is plain that, because of the constant difference of potential which is maintained between the electrodes, the ions have a definite rate of travel through the electrolyte. Accordingly the number of ions per second that pass a given cross-section of the cell remains constant, and therefore the number per second that are deposited upon the electrodes likewise remains constant. From this it is obvious that the number that are thus deposited in a given time depends upon the length of that time, or that the amount of the substance deposited upon the electrodes by the current is proportional to the time for which the current flows.

418. The Second Law. The truth of Faraday's second law may also be seen at once by remembering that the flow of current through an electrolyte means the directed movement of the ions, as well as the discharge of these ions and their deposition upon the electrodes. That is to say, electricity cannot flow through an electrolyte unless it is carried as individual charges upon the ions, and the ions cannot be deposited upon the electrodes until they are discharged. The discharge of a positively charged ion, however, necessitates the flow upon it from the cathode of one electron for each bond that thus becomes free, and the discharge of a negatively charged ion requires the flow from this ion to the anode likewise of one electron for each chemical bond.

It is thus seen that *the deposition upon each electrode of a single ion is invariably and inevitably associated with the flow through the circuit of a perfectly definite number of electrons;* and as the movement of electrons through the circuit constitutes the electric current, it is evident, in the fullest sense of the word, that the number of ions thus deposited is absolutely proportional to the strength of the current which flows.

Since in the coulomb we have a unit quantity of electricity which embodies the element of time as well as the strength of the current, we may combine Faraday's first and second laws into one, as follows:

1 and 2. The amount of substance deposited in an electrolytic cell is proportional to the number of coulombs of electricity that have passed.

419. The Flow of the Current Through Electrolytes of Different Valencies. Before it is possible to discuss Faraday's third law, it is necessary to show the effect of the current in electrolytic cells containing ions which have *different numbers of chemical bonds, or different valencies.*

All chemical elements are divided into groups or classes according to the number of chemical bonds which the atoms of these elements possess. Those with a *single bond,* such as hydrogen, sodium, silver, chlorine, are called *monads,* and are said to be *monovalent;* those *with two* bonds, such as zinc, copper, nickel, oxygen, SO_4, are called *dyads,* or are *bivalent;* those with *three bonds,* such as gold and nitrogen, are called *triads,* or are *trivalent,* etc.

If, now, a number of electrolytic cells are joined *in series,* so that the current which flows through one must flow through all, and if the several cells contain ions of different valencies, it may be shown that *the same number of chemical bonds must be set free* *in each cell.* Indeed, we may see that this must be the case if we consider that the current which flows from one cell to the next, and that therefore the current which flows throughout the entire circuit, is simply the electrons that flow from the cathode off upon the positively charged ions on one side of the electrolytic cell, and from the negatively charged ions off upon the anode on the other.

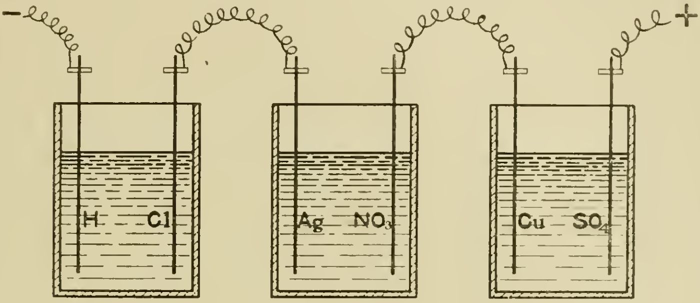

FIG. 257. Electrolytic cells in series.

For example, if there are three such electrolytic cells, Fig. 257, the first containing a solution of hydrochloric acid, HCl, the second a solution of silver nitrate, $AgNO_3$, and the third a solution of copper sulphate, $CuSO_4$, and if a current is sent through these cells in a series, H will be deposited upon the cathode and Cl upon the anode of the first cell, Ag upon the cathode and NO_3 upon the anode of the second cell, and Cu upon the cathode and SO_4 upon the anode of the third cell. That is to say, in the first cell an electron from the battery flows off from the cathode upon an H ion, thereby discharging it and depositing it upon this electrode, while at the same time a Cl ion touches the anode, gives to it the one electron with which it was charged, and it also is deposited. This electron then flows over to the cell containing the silver solution, where it is able likewise to discharge and to deposit one Ag ion, since the Ag ion also has but one bond. At the same time an NO_3 ion touches the anode, gives to it its one electron, and is itself deposited. The electron from the NO_3 ion then flows over to the copper cell, where it is able to discharge one of the two chemical bonds which the Cu ion possesses.

It is, however, impossible to discharge less than an atom of a *substance;* and so, if a Cu ion with its two bonds is to be discharged, two electrons must flow to this cell from the silver cell, and therefore two electrons must flow to the silver cell from the hydrogen cell. Accordingly, in the first cell it is necessary for two H ions and two Cl ions to be deposited. The two electrons that were in excess upon the two Cl ions then flow over to the silver cell, where they discharge and deposit two Ag ions, while two NO_3 ions are discharged and deposited upon the anode of this cell. The two electrons from the two NO_3 ions flow over to the copper cell, where they are able to discharge one ion of Cu, with its two bonds, and with its consequent deficit of two electrons. This atom of copper is then deposited upon its cathode, while one SO_4 ion yields its two electrons to the anode of this cell, and is deposited upon it. The two electrons thus released from the SO_4 ion then flow back to the battery, thus constituting the current which flows during this process throughout the circuit.

Since the presence or the absence of an electron upon any atom protects one chemical bond of that atom, and since the current which flows from one cell to the next is simply the electrons that have been released from the ions of the previous cell, it is evident that the same number of electrons must come from one cell as from any other, and therefore that *the number of bonds released, or set free, must be the same in each cell.*

420. The Third Law. Faraday's third law has to deal not only with the valencies of the various atoms, but with the relative *weights* of these atoms with respect to hydrogen, which is taken as unity. If we continually bear in mind the fact that *the atoms of the various elements have different weights,* it will be evident that the third law must follow as a necessary consequence.

In the previous paragraph, we have seen that, for each atom of Cu deposited, two atoms of H and two atoms of Ag must likewise be deposited. The actual *masses* of these substances, however, vary widely from this simple relation, inasmuch as the atoms of the various elements have different weights. Although we are ignorant of the absolute weights of the atoms, we do know, nevertheless, their weights with relation to one another. Relative to the weight of an atom of hydrogen, which is taken as 1, the weight of an atom of silver is 108, and the weight of an atom of copper is 63.4. Accordingly, the relative *weights* of the above two atoms of H, two atoms of Ag, and one atom of Cu, are in the ratio, 2 : 216 : 63.4, or 1 : 108 : 31.7; that is, for every 1 gram of H deposited, 108 grams of Ag, and 31.7 grams of Cu will likewise be deposited. These quantities, namely, 1 gram of H, 108 grams of Ag, 31.7 grams of Cu, are said to be *chemically equivalent* to one another; and these are the relative amounts of the various substances that by Faraday's third law will be deposited in the several cells.

421. The Chemical Equivalent of a Substance. *By the chemical equivalent of a substance is meant the number of grams of that substance which can replace 1 gram of H in any chemical compound.* Since the atom of Ag and the atom of H are each monovalent, one *atom* of Ag will everywhere replace one *atom* of H; but as the atom of Ag is 108 times as heavy as the atom of H, the number of *grams* of Ag in any compound is 108 times as great as the number of *grams* of the H which it replaces. Therefore, 108 grams of Ag and 1 gram of H are chemically equivalent.

In a similar manner, since the atom of Cu is bivalent, *one atom* of Cu will replace *two atoms* of H in any chemical compound; but, as the atom of Cu has a relative weight of 63.4, while the two atoms of H have a relative weight of 2, the number of *grams* of Cu in any chemical compound is 63.4/2, or 31.7 times as great as the number of *grams* of H which it replaces. Therefore, 31.7 grams of Cu are chemically equivalent to 1 gram of H; or, the chemical equivalent of Cu is 31.7. From this may be drawn the important deduction that *quantities of substances that are chemically equivalent are also electrically equivalent.*

An examination of the above reasoning will show that *the chemical equivalent of a substance is always the ratio of the atomic weight to the valency.*

422. The Electrochemical Equivalent of a Substance. Another term quite similar to the chemical equivalent, but of different meaning, namely, *the electrochemical equivalent*, is of frequent occurrence in this connection. *The electrochemical equivalent of a substance is the actual number of grams of that substance which will be deposited by 1 coulomb of electricity.*

Thus, by the definition, we see that the electrochemical equivalent of Ag is 0.0011180. So also, the electrochemical equivalent of Cu is 0.0003281 and that of H is 0.00001036.

This value of the electrochemical equivalent of H gives us some idea of the enormous quantities of electricity which ions carry as their ionic charges. Since 1 coulomb of electricity will set free 0.00001036 grams of H, we see that to set free 1 gram, or 11.16 litres, of H requires $\frac{1}{0.00001036}$ or 96,530 coulombs. Therefore, a single gram of H in the ionic state carries a quantity of electricity equal to 96,530 coulombs. This is also the number of coulombs required to liberate by electrolysis the mass of one chemical equivalent of any other substance. Moreover, when we remember that 1 coulomb is equal to 3,000,000,000 electrostatic units of electricity, we see clearly why even a large static machine is able to produce only the feeblest electrolytic decomposition. To decompose and set free a single gram of H by electrostatic means would require the expenditure of 96,530 × 3,000,000,000, or 289,000,000,000,000 electrostatic units, which is possibly more electricity than is developed during a severe thunder storm.

423. The Voltameter. The knowledge of Faraday's laws, together with the correct determination of the electrochemical equivalents of silver and copper, enables us most accurately to measure a current by weighing the amount of metal deposited in a certain time by the current in an electrolytic cell. The cell for accomplishing this purpose, which is known as the *voltameter,* is provided with a cathode which may be taken out and weighed before the current flows, and weighed again after the metal has been deposited. Of the two voltameters, the silver and the copper, which are in frequent use, the silver voltameter is the more accurate owing to the higher chemical equivalent of silver, although the copper voltameter is the more practical, because of its greater cheapness and greater ease of manipulation.

In the copper voltameter, copper plates are used as the electrodes, and are immersed in a solution of $CuSO_4$. The cathode is carefully weighed before being placed in the electrolyte; and after the current is allowed to flow for a measured interval of time, it is taken out, washed, dried, and weighed again, the difference in weight being the amount of copper deposited during the time the current passed. Since the electrochemical equivalent of Cu is known, the number of coulombs which passed through the electrolyte may then be accurately determined. If the strength of the current has remained constant during the time, this number of coulombs may be divided by the number of seconds during which the current flowed, and the strength of the current in amperes thus obtained. If the current has been of varying strength, the voltameter will still give the total number of coulombs that have passed, and therefore the average current that has flowed during that time.

A consideration of the processes in electrolysis will show that for every ion of Cu deposited upon the cathode, an ion of SO_4 is deposited upon the anode. If the anode had been of platinum. the SO_4 would have broken up, setting free an atom of O, and uniting with a molecule of water to form a molecule H_2SO_4, as in the former examples of electrolysis. In the copper voltameter, since the positive plate is likewise of copper, the SO_4 ion, when set free at this electrode, at once attacks an atom of Cu upon the anode, thus forming a molecule of $CuSO_4$, which, of course dissociates into Cu ions and SO_4 ions.

In the copper voltameter, therefore, for every atom of Cu that is deposited upon the cathode, an atom of Cu is taken from the copper anode, and the number of Cu and SO_4 ions remains nearly constant, no matter for how long a period the electrolysis may be continued. It might seem, therefore, that the number of coulombs which have passed could be determined from the loss in weight of the anode as well as from the gain in weight of the cathode, and, indeed, this is approximately true. It is found. however, that, possibly owing to slight impurities in the copper. or for other reasons, the loss on the part of the anode is not quite as regular as is the gain on the cathode, and so it is only the cathode which is weighed when current is determined by means of the voltameter.

In the silver voltameter, the anode is a rod of silver. while the cathode is usually a platinum crucible or a hollow, bowl-shaped platinum vessel, upon the interior of which the deposition of silver takes place. A vessel of this form is used instead of a plate, since the silver is deposited as a multitude of tiny crystals, many of which might easily be lost from the surface of a plate during the process of washing. The electrolyte used in this voltameter is a solution of $AgNO_3$. As the Ag is deposited upon the cathode, the NO_3 which is set free at the anode, attacks an atom of Ag upon the silver rod, forming again a molecule of $AgNO_3$ which dissociates at once into Ag and NO_3 ions, thereby maintaining constant the strength of the electrolyte. Thus the anode theoretically loses as much as the cathode gains, although, for the same reason as before, it is only the gain on the cathode that is considered in determining the current.

Since the chemical equivalent of silver is 108, while that of copper is but 31.7, it follows that if a silver and a copper voltameter were placed in series, a current which would deposit 108 grams of silver would deposit only 31.5 grams of copper in the same time, therefore, because of the very high chemical equivalent of silver, the silver voltameter is the more reliable.

424. Electroplating and Electrotyping. This process of electrolysis is employed in the arts in electroplating and electrotyping. If an object is to be silver plated, it is suspended so as to form the cathode in an electrolytic cell, while a plate of silver is hung opposite as the anode, the electrolyte being a solution of the cyanide of silver. The current in passing through the electrolyte deposits a coating of silver upon the object to be plated, while the anode loses in weight. In a similar manner, an object to be copper plated is made the cathode, a plate of copper is suspended opposite as the anode, and the electrolyte is $CuSO_4$.

In electrotyping, a mould in wax or plaster is made of the type or object to be reproduced, and this, after being coated with black lead to make it a conductor, becomes the cathode of an electrolytic cell, a copper plate opposite being the anode, and the electrolyte in this case also is $CuSO_4$. The passage of current through the solution causes a layer of copper to be deposited upon the wax mould, conforming in the minutest detail to every line or impression of the mould. When this layer becomes sufficiently thick, it may be stripped off and reenforced by type metal poured upon the back, thereby furnishing a firm, hard, enduring copper surface from which printing may be done.

425. Advantage Taken of the Counter Electromotive Force. If platinum plates are used as electrodes in a solution of H_2SO_4, we have seen (p. 478) that oxygen will be evolved at the anode and hydrogen at the cathode. This collection of gases at the electrodes developes a counter electromotive force in the cell, since these gases tend again to dissolve in the electrolyte and form ions. Indeed, in every electrolytic cell there is thus developed a counter electromotive force due to polarization, except in a cell like the copper or silver voltameter, where the anode consists of a plate of the same metal as that which is being deposited upon the cathode, in which case there is no counter electromotive force. Advantage, however, is taken of the counter electromotive force which is developed in ordinary electrolytic cells to make what is known as secondary batteries, or accumulators, or, more commonly, storage batteries, although, as will be seen, such a battery does not accumulate or store up electricity.

426. The Accumulator, or the Storage Battery. If a plate of lead dioxide, PbO_2, and a plate of metallic lead, Pb, are placed in an electrolyte of dilute H_2SO_4, they at once assume different potentials, the former being about two volts higher than the latter. These two plates then correspond to the copper and the zinc plates in a primary battery; so that if they are connected by a wire, a current will flow from the high potential PbO_2 plate through the wire to the low potential Pb plate.

By this flow, phenomena similar to those which appear in the simple battery now take place, two H ions are discharged and deposited on the PbO_2 plate, just as H ions appear upon the copper plate of the simple battery, while in a similar way, an SO_4 ion is deposited upon the low potential Pb plate. At the positive plate, a secondary reaction then takes place between the PbO_2, the released H_2, and a molecule of H_2SO_4, according to the following reaction:

$$PbO_2 + H_2 + H_2SO_4 = PbSO_4 + 2H_2O.$$

At the negative plate the SO_4, which is there released, combines directly with the lead, forming lead sulphate, $PbSO_4$, according to the reaction.

$$Pb + SO_4 = PbSO_4.$$

By the flow of the current, therefore, ions are liberated at both plates; and the secondary reactions which then take place change the material of these plates into lead sulphate, $PbSO_4$. The two plates have thus become alike; and, as it is impossible to obtain a difference of potential between two similar plates in a single electrolyte, the flow of current ceases, and the battery is said to be discharged.

These two plates of $PbSO_4$ may now be regarded as the anode and the cathode of an electrolytic cell, and a current from an outside source may be sent through this cell. If the anode of this electrolytic cell is then the plate which was formerly the PbO_2 plate, and if the cathode is that which was formerly the Pb plate, SO_4 ions will be discharged upon the anode and H_2 ions upon the cathode, as in the ordinary electrolysis of dilute H_2SO_4. At the anode the SO_4 will combine with the $PbSO_4$ of the plate and with two molecules of water, according to the reaction.

$$PbSO_4 + SO_4 + 2H_2O = PbO_2 + 2H_2SO_4,$$

while at the cathode the H_2 will reduce the plate of $PbSO_4$ to metallic lead, by the reaction.

$$PbSO_4 + H_2 = Pb + H_2SO_4.$$

Thus, by regarding the discharged battery as an electrolytic cell, and sending through it a current from an outside source, ions are released at the electrodes; and the secondary reactions which then take place change the plates of $PbSO_4$ once more into plates of PbO_2 and Pb, thereby restoring the battery to its former condition, or "charging" it. The battery thus "charged" may now again be used as a primary battery until it becomes discharged after which it is necessary once more to send a current through it, or to charge it, and so on.

It is to be noticed in this connection that the storage battery stores up energy but not electricity. In other words, the energy of the current which is sent into the battery is expended in producing chemical changes upon the electrodes, and therefore remains as potential energy until a wire connects the two plates, when it again appears as the energy of the electric current. In a good storage battery the electrical energy that has been expended in charging the battery may remain for weeks as potential energy without serious loss.

It is only fair to say that there are other secondary reactions which take place both in the charging and discharging of a storage battery, some of which, indeed, are not well understood. The simple one here given, however, is undoubtedly the one to which the action of the storage battery or the accumulator is principally due.

CHAPTER XXXIV.

ELECTRICAL UNITS.

427. The Identity of Static and Dynamic Electricity. In a previous section (p. 488), it was shown that the quantity of electricity which is developed by batteries is many times greater than that which is ordinarily produced by electrostatic means. All of the results, however, which may be obtained from the flow of the electric current from a battery or other source of difference of potential may likewise be reproduced by the use of the Holtz machine, only on a very much smaller scale. Thus, it is perfectly possible to use the current from a static machine for the electrolysis of water; but the *quantity* of electricity which flows is so small that the production of hydrogen and oxygen may be observed only by concentrating the evolution of these gases upon tiny electrodes not more than a fraction of a square millimeter in area, and operating the machine for a long time. So, also, if the current from this machine is sent through a delicate galvanometer, which is a device for detecting the flow of current, it will produce only the feeblest effect, while the current from even the smallest battery will disturb the instrument violently.

428. The Electromagnetic Unit of Quantity. From the foregoing paragraph we see that, in the study of dynamic or current electricity, we have to consider quantities of electricity which are enormously greater than those obtained from electrostatic sources. In order, therefore, to avoid the constant use of exceedingly large numbers, a unit of quantity many times larger than the electrostatic unit has been adopted, and this, like the unit of current, is founded upon an entirely different system of units, known as the electromagnetic system.

It will be shown later (p. 541) that the strength of an electric current may be expressed in terms of a unit which depends only upon the magnetic relations of the current; and on this account the unit of current so determined is known as the electromagnetic unit. Moreover, since the unit of current enters into the definition of all other electrical units, the entire system of units thus based upon the electromagnetic unit of current is called the electromagnetic system. We may therefore make use of this electromagnetic unit of current, of which the ampere is the 1/10th part, as if it had already been described, reserving, however, the strict definition of this unit until we have made a detailed study of the flow of the current and of its magnetic relations.

In a previous section (p. 350) it was shown that a flow of current is interpreted as meaning a directed movement of the electrons, and therefore as the transport of a certain amount or quantity of electricity. Moreover, when the current is unvarying in strength, the quantity of electricity which passes any cross-section of a conductor is the same as that which passes any other cross-section, so that *a current of the same strength or intensity flows through every part of the circuit.* If either the strength of the current or the time of flow were increased, the quantity of electricity that passes would likewise be increased. From this it follows that the quantity Q of electricity transported by a current is proportional to the intensity I of the current and to the time t for which this current flows: that is,

$$Q : I t,$$

or

$$Q = k\,I\,t,$$

where k is the factor of proportionality. Inasmuch as the intensity I of the current is determined independent of all other electrical quantities, we may define the unit of quantity in terms of the unit of current and the unit of time. In order, therefore, that the factor of proportionality may be unity, we agree to call the quantity unity when the strength of the current and the time are each equal to unity, in which case

$$k = 1,$$

and

$$Q = I\,t.$$

Accordingly, *the unit of quantity in the electromagnetic system is defined as the quantity of electricity that passes through a circuit when 1 electromagnetic unit of current flows for 1 second of time.* By comparing the electromagnetic and the electrostatic units of quantity, it is found that 1 electromagnetic unit of quantity equals 30,000,000,000 or 3×10^{10} electrostatic units.

429. The Coulomb. Because the electromagnetic unit of current is relatively a very strong current (p. 482), it is usual in practice to take the 1/10th part of this current as a practical or working unit, to which the name of *the ampere* has been given. From this it follows that only 1/10th of an electromagnetic unit of quantity passes when a current of 1 ampere flows for 1 second, and this is know as the *coulomb*. *The coulomb, which is the practical or working unit of quantity, is thus defined as the quantity of electricity that passes through a circuit when 1 ampere flows for 1 second.* It is thus seen that the coulomb is the 1/10th part of the electromagnetic unit of quantity, and is equal to 3,000,000,000, or 3×10^{9} electrostatic units.

430. The Energy of the Electric Current. In the treatment of electrostatics, we have defined the potential at a point (p. 356) as the potential energy which the unit quantity of positive electricity will possess when brought to that point. Moreover, we have seen that the difference of potential between two points is measured by the number of ergs of work that must be expended in carrying the unit of positive electricity from one point to the other. Both of these conceptions find their application in dynamic or current electricity; but, as here the unit quantity of electricity is expressed in the electromagnetic system, so also must the unit of potential likewise be expressed in this system.

In order to determine the magnitude of this unit, we consider that a difference of potential of E units exists between two points, A and B, Fig. 258, where A is higher in potential than B. Since work is performed in carrying a unit of electricity from a low to a high potential, it follows that if 1 unit of electricity is carried from B to A, an amount of work will be performed which is proportional to the difference of potential E, and if Q units are thus carried, an amount of work will be performed which is proportional to QE. On the other hand, if a current flows from A to B, whereby Q units of electricity move of themselves through the difference of potential E, *these Q units will then be able to perform an amount of work which is proportional to QE*. But (p. 495), since

FIG. 258. The points A and B are at different potentials.

$$Q = It,$$

it is evident that the passage of Q units of electricity from A to B is equivalent to a current of intensity I flowing for t seconds between these points, and that this current will then perform an amount of work proportional to IEt.

That is,

the energy of the current: $I\,E\,t$,

or

$$\text{the energy of the current} = k\,I\,E\,t,$$

where k is the factor of proportionality.

If, for the sake of simplicity, we now agree that the difference of potential shall be called unity when the current, the time, and the energy are each equal to unity, then

$$k = 1,$$

and the above equation becomes

$$\text{the energy of the current} = I\,E\,t.$$

At a later time (p. 517) an equivalent and alternative expression for the energy of the current will be given, which will be found of frequent service.

431. The Electromagnetic Unit of Electromotive Force or Difference of Potential. In order that k may have the value of unity in the expression for the energy of the current, we have been obliged to limit the value of the unit difference of potential to the above conditions. These limitations therefore specify the value of this unit, which may, accordingly, be defined as follows: *The electromagnetic unit of electromotive force or difference of potential is of such magnitude that it will enable the electromagnetic unit of current in 1 second of time to perform 1 erg of work.*

432. The Practical or Working Unit of Electromotive Force —The Volt. When we remember that the electromagnetic unit of current is relatively a strong current, and that the erg is a very small unit, we see that the electromagnetic unit of difference of potential, as here defined, must be exceedingly small. It is so small, indeed, that a multiple by 100,000,000, or 10^8, is taken arbitrarily to obtain a working or practical unit, and this is called the *volt*. Accordingly, *the volt, which is the practical or working unit of electromotive force, is equal to* 10^8 *electromagnetic units.*

This particular multiple was taken in order that the practical unit might be an exact multiple by 10 of the electromagnetic unit, and still represent as nearly as possible the electromotive force of the Daniell battery. Although the volt is thus an enormous multiple of the electromagnetic unit, it is, nevertheless, still only 1/300 of the electrostatic unit.

433. The Electromagnetic Unit of Capacity. If a conductor is connected with a source of high potential, such as a battery, electricity will flow upon this conductor until the potential of the latter is as high as that of the source. In this process, a perfectly definite quantity Q of electricity is required to raise the conductor to the potential E of the battery. If the conductor is now discharged and connected with a battery of higher electromotive force E', a quantity of electricity correspondingly greater, namely Q', will be required to raise the conductor to this higher potential. Therefore, the *ratio* of this greater quantity to the higher potential is exactly the same as that of the smaller quantity to the lower potential, *since the potential to which a conductor is raised is always proportional to the quantity with which it is charged.*

This constant ratio between the quantity of electricity with which a conductor is charged and the potential to which the conductor is thereby raised is thus seen to be *a characteristic of the conductor alone, and is known as its capacity.* Accordingly, if C represents this constant ratio, or the *capacity* of the conductor,

$$C = \frac{Q}{E}.$$

If E is unity, then

$$C = Q,$$

from which *the capacity of a conductor is defined as the quantity of electricity which will raise that conductor to unit poten-*

tial. If, in addition, Q is unity, then C is likewise unity; hence *a conductor has unit capacity when unit quantity of electricity will raise that conductor to unit potential.* These definitions are obviously true in any system. Accordingly, in the electromagnetic system, *a conductor has 1 electromagnetic unit of capacity when 1 electromagnetic unit of quantity of electricity will raise that conductor to 1 electromagnetic unit of potential.*

But the electromagnetic unit of capacity, as thus defined, is so enormously large that in practice it has been found advisable to adopt a unit which is only $\frac{1}{1,000,000,000}$ or 10^{-9}, of this electromagnetic unit, and this, after the great English physicist Faraday, is called the *farad.* In other words, *the farad is the capacity of a conductor or condenser which 1 coulomb of electricity will raise to the potential of 1 volt,* so that, in general,

$$\text{Farads} = \frac{\text{Coulombs}}{\text{Volts}}.$$

434. The Microfarad. Even the farad, however, is a capacity which is exceedingly large as compared with the capacities which we ordinarily have to measure. Indeed, the earth, regarded as an insulated spherical conductor, has a capacity of but 7/10,000 of a farad. In other words, the quantity of electricity represented by 1 coulomb would raise 1429 insulated spherical conductors of the size of the earth to the potential of 1 volt. In order, therefore, to obtain a still smaller unit for ordinary laboratory measurements, it has been found convenient to adopt as the working unit of capacity the one-millionth part of a farad, and this is called the *microfarad.* From this it is seen that

$$1 \text{ Microfarad} = 10^{-6} \text{ Farad},$$

$$= 10^{-15} \text{ electromagnetic unit of capacity}$$

All ordinary measurements of capacity, are, therefore, made in microfarads; for instance, the capacity of the Atlantic Cable is about one-third of a microfarad per mile. For laboratory purposes, a microfarad, with multiples and subdivisions, is usually built in the form of a condenser, Fig. 259, consisting of sheets of tin-foil separated by plates of mica. So great is the capacity of a condenser constructed in this way, that even a comparatively small box may thus contain a condenser having a capacity as large as a microfarad.

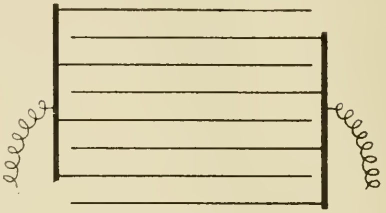

Fig. 259. The construction of a condenser.

CHAPTER XXXV.

ELECTRICAL RESISTANCE.

435. Nature of Resistance. It is found that electricity does not flow with perfect freedom through even the best conductors. Indeed, all conductors offer, in a greater or less degree, a certain opposition to the passage of the electric current, and this opposition is called *resistance*. *Electrical resistance,* therefore, *is the opposition which electricity meets in passing through a conductor.* The nature of resistance, which may possibly be somewhat analogous to friction, is not well understood. Apparently it occurs because of the difficulty of movement of the electrons between the molecules or through the molecules of a conductor, or to the opposition to motion which the electrons meet as they pass from molecule to molecule through the material of which the conductor is made.

436. Resistance of Conductors. Since electrical resistance is a subject of great importance in the flow of electricity, it is necessary definitely to determine the quantities upon which it depends. In the first place, it is found by experiment that *the resistance of a conductor is directly proportional to its length;* that is, if a conductor of a certain length offers a certain resistance, one in every way similar to the first but twice as long, will offer twice as much resistance. It is also found, other things being equal, that the larger the conductor, the smaller is the resistance, or that *the resistance of a conductor is inversely proportional to the area of its cross-section.* Moreover, it is found that *the resistance of a conductor varies with the material of which the conductor is made,* so that a conductor of copper, for instance, offers a certain resistance, while one of iron of the same length and cross-section offers an entirely different resistance.

Accordingly, since the resistance, R, of a conductor is directly proportional to the length, L, and inversely proportional to the cross-section, A, of the conductor, we see that

$$R : \frac{L}{A},$$

or

$$R = k\frac{L}{A},$$

where k is the factor of proportionality.

437. Specific Resistance. In the above expression, the factor k, which has different values for different materials, is, therefore, the quantity which expresses the dependence of the resistance upon the material of which the conductor is made, and is called the *specific resistance of the material.* From this it is obvious that if L and A are each equal to unity, then

$$R = k.$$

In other words, *the specific resistance of a material is the actual resistance of a conductor of that material 1 centimeter long and 1 square centimeter in cross-section.*

438. Ohm's Law. We are now in a position to discuss the flow of the electric current, and to show the relation that exists between the quantities upon which this flow depends. As regards the influence of the difference of potential, it is proved by experiment that *the intensity I, of the current that flows is directly proportioned to the difference of potential, E, which causes it to flow.* That is, if a certain difference of potential causes a current of a certain intensity to flow, other things being equal a difference of potential twice as great will cause a current of twice this intensity to flow. Again, as regards the influence of the resistance, it is proved by experiment that *the intensity of the current that flows is inversely proportional to the resistance through which it flows.* Thus, if a certain resistance will permit a current of a certain intensity to flow, other things being equal a resistance of twice this magnitude will permit a current of only one-half this intensity to flow. That is, if the resistance is doubled, the current is diminished by half, and so on. Accordingly, since the intensity, I, of the current that flows is directly proportional to the difference of potential, E, which causes it to flow, and is inversely proportional to the resistance, R, which impedes its flow, it follows that

$$I : \frac{E}{R}$$

or

$$I = k\frac{E}{R},$$

where k is the factor of proportionality.

If we agree to call the resistance unity when the current and the electromotive force are each equal to unity, then

$$k = 1,$$

and the equation becomes

$$I = \frac{E}{R}$$

This equation, which holds true for all constant currents, expresses the relation between the current, the electromotive force, and the resistance, and is known as *Ohm's law*. *By Ohm's law, therefore, the intensity of the current that flows is equal to the ratio of the electromotive force which causes it to flow to the resistance through which it flows.*

439. The Electromagnetic Unit of Resistance. In thus placing $k = 1$ in the above equation, we have, however, thereby defined the unit of resistance to be the resistance through which the unit of electromotive force will cause the unit of current to flow. Inasmuch as the relation which expresses Ohm's law is true for any system of units, it must also be true for the electromagnetic system; and since the units of electromotive force and current in this system are known, we may then define the unit of resistance in terms of these units. Accordingly, *the electromagnetic unit of resistance is the resistance through which 1 electromagnetic unit of electromotive force will cause 1 electromagnetic unit of current to flow.*

It is to be observed that the unit of resistance, like the unit of electromotive force, depends for its magnitude upon the electromagnetic unit of current, the definition of which, for reasons given above (p. 494), must be deferred until another time (p. 541).

440. The Practical or Working Unit of Resistance—The Ohm. If however, the *practical* or *working* units of current and electromotive force, instead of the electromagnetic units, are used in Ohm's law, that is, if the unit of current is taken as 1 ampere and the unit of electromotive force as 1 volt, then the unit of resistance through which this current flows is the practical or working unit of resistance, and is called the ohm. Accordingly, *the ohm, which is the practical or working unit of resistance, is defined as the resistance through which an electromotive force of 1 volt will cause a current of 1 ampere to flow.* That is,

$$1 \text{ Ampere} = \frac{1 \text{ Volt}}{1 \text{ Ohm}},$$

or, in general

$$\text{Current (in amperes)} = \frac{\text{Electromotive force (in volts)}}{\text{Resistance (in ohms)}}.$$

The relative magnitudes of the ohm and the electromagnetic unit of resistance may easily be determined from the relations in Ohm's law. Since 1 ampere = 10^{-1} electromagnetic unit of current, and 1 volt = 10^{8} electromagnetic units of electromotive force, we see that the equation

$$1 \text{ Ampere} = \frac{1 \text{ Volt}}{1 \text{ Ohm}},$$

when expressed in electromagnetic units becomes

$$10^{-1} = \frac{10^{8}}{10^{9}},$$

That is, since the left hand side of this equation has the value 10^{-1}, the right hand side must have this same value; and as the numerator of the fraction is 10^{8}, the denominator, which expresses the number of electromagnetic units of resistance in the ohm, must be taken as 10^{9}. Accordingly, in order that the relations in Ohm's law may hold, the ohm *must* be taken as equivalent to 10^{9} electromagnetic units of resistance.

The selection of the number of electromagnetic units of resistance in the ohm is therefore not a free choice, but is definitely prescribed by Ohm's law. We have chosen arbitrarily the fractional part of the electromagnetic unit of current which we call the ampere, and we have chosen, also arbitrarily, the number of electromagnetic units of electromotive force in the volt. Having done this, we are then *obliged* to take the ohm as the equivalent of 10^9 electromagnetic units of resistance, in order that the relations in Ohm's law may hold true.

441. The International Ohm. Although the ohm is defined as the resistance through which an electromotive force of 1 volt will cause a current of 1 ampere to flow, its practical measurement in the laboratory would be difficult without further aid. Accordingly, there has been adopted a standard of resistance, called the international ohm, which *represents* the resistance of an ohm, as above defined, within the errors of modern electrical measurements. This standard of resistance is specified as follows:—"The international ohm is the resistance offered to an unvarying electric current by a column of mercury at the temperature of melting ice, 14.4521 grams in mass, of a constant cross-sectional area, and of a length 106.3 centimeters." The cross-section of the column of mercury thus specified is very nearly 1 square millimeter in area.

The practical method of making a measurement of resistance is to compare the resistance to be determined with the international ohm, as here defined, or with a resistance which itself has been made by comparison with this international standard.

442. The Resistance Box, or Rheostat. While it is thus possible at any time to reproduce the international ohm, it is impracticable actually to use a unit so fragile and so difficult of management as this mercury column. Accordingly, wires have been adjusted in length until they offer an electrical resistance identical with the resistance of this column of mercury, and these are the secondary standards that are practically used. Moreover, by taking wires of suitable lengths, multiples and subdivisions of the ohm may readily be made; and a number of these resistances are generally grouped together in a box, forming what is known as a resistance box, or a rheostat.

The top of such a resistance box is usually provided with a series of brass blocks, between which brass plugs may be made to fit tightly. Underneath is the wire, the resistance of which is to be used, one end being joined to one block, Fig. 260, and the other end being connected to the block adjacent. To prevent self-induction, this wire is wound on the spool half in one way and half in the other, that is, it is doubled back upon itself. When the plug is inserted between the blocks, the current flows from one block through the plug to the next; but when the plug is withdrawn, the current, in passing from block to block, must pass through the wire, the resistance of which is to be introduced. By having a number of different coils in such a box, any desired resistance may be included in the circuit by pulling out the apropriate plugs. A box with no plugs withdrawn has practically a zero resistance.

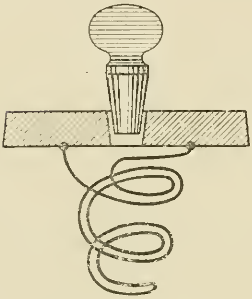

Fig. 260. A resistance coil.

443. Variation of Resistance with Temperature. The resistances of all bodies vary with a change of temperature. Metallic conductors, in general, increase in resistance as the temperature rises, and decreases in resistance as the temperature falls. Indeed, through wide thermal variations, the resistances of the pure metals are almost exactly proportional to their absolute temperature. This for a long time gave rise to the erroneous belief that the resistances of the pure metals would become zero if the temperature could be reduced to the absolute zero. As this point is approached, however, deviations from this strict proportionality occur, which show that the resistances of the pure metals are still finite, though very small, even at the absolute zero.

Two exceptions are found to the general law that a rise of temperature produces an increase in the resistance of a conductor, namely: carbon and electrolytes, and with these the resistance becomes less as the temperature rises. The resistance of the carbon filament of a 16 c. p. incandescent lamp, which is about 400 ohms when cold, falls to about 200 ohms when the lamp is heated to a bright incandescence.

It is perhaps, not difficult to understand why it should be true that the resistance of electrolytes becomes less at higher temperatures. In this connection, we must remember that the conduction of electricity by an electrolyte means the carrying of individual electric charges upon the ions, and that the kinetic energy of these ions increases as the temperature becomes higher. We can see, therefore, how the movement of the ions is facilitated by the increased kinetic energy of the liquid particles, that is, by the higher temperature of the electrolyte.

The decrease in the resistance of the metals with a fall of temperature may perhaps likewise be explained as a result of the diminished molecular kinetic energy. If the resistance of a metallic conductor means the difficulty of the passage of the free electrons from molecule to molecule in this conductor, it is possible to understand how this difficulty might become less if the molecules were moving less violently, that is, if the temperature of the conductor were to fall. A *decrease in the resistance,* however, means an *increase in the conducting power, or the conductivity,* of a substance; hence, *the conductivity of the metals increases as the temperature becomes lower,* that is, the metals become better conductors as the kinetic energy of the molecules becomes less.

That the electrons have much to do, not only with the conduction of electricity by metallic conductors, but also with the conduction of heat by metals, (p. 249), receives strong confirmation from the fact that the relative conductivities of the various pure metals for electricity are arranged in the same order as are the conductivities of these same metals for heat.

444. The Fall of Potential Along a Wire Carrying a Current. Let us suppose that the current I is made to flow through the two resistances r_1 and r_2 placed in series. Let e_1 be the difference of potential between the two ends of the resistance r_1, Fig. 261, and e_2 the difference of potential between the two ends of r_2. Since the same current I must flow through both resistances, we have, by Ohm's

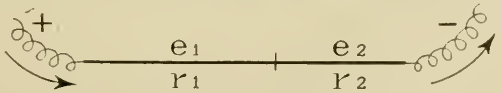

FIG. 261. The fall of potential along a conductor is proportional to the resistance.

law,

$$I = \frac{e_1}{r_1} = \frac{e_2}{r_2},$$

or

$$\frac{e_1}{e_2} = \frac{r_1}{r_2}.$$

This shows that when a current flows through several conductors in series, a difference of potential exists between the ends of these conductors proportional to the resistance of each. In other words, *when a wire carries a current, a fall of potential occurs along that wire proportional to the resistance through which the current flows.*

445. The Divided Circuit. When a circuit divides so that two or more branches are offered to the passage of the current, the current also divides, There are, therefore, as many divisions of the current as there are branches of the conductor. If the resistances of these branches are all equal, the current divides equally between the several paths thus offered; but if the branches are of unequal resistance, the flow of current through any branch will be inversely as the resistance of that branch.

For example, if a circuit divides into two branches, the relative resistances of which are 2 and 3 respectively, Fig. 262, the currents I_1 and I_2 in these branches will be proportional to the reciprocals of the resistances, since the difference of potential between the two ends of each branch is the same. Accordingly, if e is the difference of potential between the two ends of the resistances,

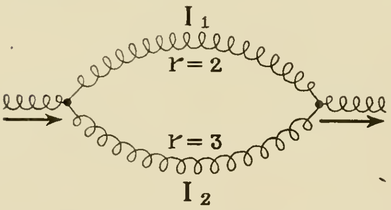

FIG. 262. The current divides inversely as the resistance.

$$I_1 : I_2 = \frac{e}{2} : \frac{e}{3},$$

or

$$I_1 : I_2 = \frac{1}{2} : \frac{1}{3},$$

that is,

$$I_1 : I_2 = 3 : 2.$$

Therefore, in the branches in which the relative resistances are as 2 : 3, the currents will be as 3 : 2. In other words, in this example, 3/5 of the current will flow through the branch having the resistance 2, and 2/5 will flow through the branch having the resistance 3.

When a circuit divides, the different branches of the circuit are said to be in multiple, or in parallel, with one another. Each branch of a divided circuit is called a *shunt,* so that when a number of conductors are in multiple, each branch forms a shunt to the others.

446. The Resistance of Conductors in Multiple, or in Parallel. We have seen that the resistance of a conductor is the opposition which the current meets in passing through the conductor. *The less the resistance,* the greater is the freedom with which the current flows and *the greater is the conductivity. Conductivity, therefore, is the reciprocal of the resistance.*

If a circuit divides into two branches, the resistances of which and r_1 and r_2, Fig. 263, the current I also divides into currents I_1 and I_2, which flow through the resistance r_1 and r_2 respectively. Since there can be no accumulation of current at any point of branching, the total current in the branches must equal the original current before dividing, or

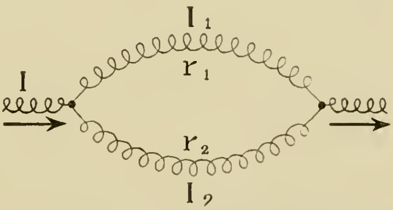

FIG. 263. Conductors in multiple, or in parallel.

$$I = I_1 + I_2. \tag{1}$$

If e denotes the difference of potential between the two ends of the divided circuit, we have, by Ohm's law (p. 502),

$$I_1 = \frac{e}{r_1},$$

and

$$I_2 = \frac{e}{r_2}.$$

If r denotes the *equivalent* resistence of the two branches which are in multiple, that is, the single resistance that could replace the resistances r_1 and r_2, we have, since I is the total current flowing,

$$I = \frac{e}{r}.$$

Therefore, from (1),

$$\frac{e}{r} = \frac{e}{r_1} + \frac{e}{r_2},$$

or

$$\frac{1}{r} = \frac{1}{r_1} + \frac{1}{r_2}.$$

That is, *the conductivity of two wires in multiple is the sum of the separate conductivities of these wires.* Clearing this equation of fractions,

$$r_1 r_2 = r r_2 + r r_1,$$

from which

$$r = \frac{r_1 r_2}{r_1 + r_2}.$$

This expression gives the resistance of two wires in multiple or parallel where r_1 and r_2 are separate resistances.

In a similar manner, if three conductors, r_1, r_2, r_3 are joined in multiple, the resistance r of these three is

$$r = \frac{r_1 r_2 r_3}{r_1 r_2 + r_1 r_3 + r_2 r_3}.$$

If conductors are placed in series, the total resistance is simply equal to their sum.

447. Resistance of Batteries. Like all other conductors, electrolytes offer resistance to the flow of current; and therefore the resistance which the current meets in passing through the fluids of a battery may be a very important factor in determining the flow of current. Indeed, the resistance offered by the electrolyte of a battery is perhaps twenty million times as great as that of a copper conductor of the same cross-section.

While the resistance in a battery may be diminished by using large plates, so that the cross-section of the liquid conductor may be large, the same result may be attained by using several battery cells and connecting them in multiple or parallel. If two exactly similar cells are so joined, the resistance of the two in multiple (p. 510) will be $\frac{r}{2}$, where r is the resistance of each cell; and if n cells are joined in multiple, the total resistance is $\frac{r}{n}$.

If, however, n similar cells are joined in series, the resistance will be nr, since this is equivalent to increasing the length of the liquid conductor.

448. Batteries Connected in Series. Batteries should be joined in series or in multiple, according to the purpose for which they are used, as may be seen from the following consideration. Let e be the electromotive force of a single cell, r the internal resistance of a single cell, R the resistance outside the battery, or the external resistance, and n the number of battery cells. By Ohm's law (p. 502), the current, I, produced by a single cell flowing through the external resistance R is

$$I = \frac{e}{r + R},$$

and if n cells are joined *in series,* the current is

$$I = \frac{n e}{nr + R}.$$

That is, the electromotive force of the battery is increased in proportion to the number of cells (p. 474), but the internal resistance of the battery is likewise increased in the same proportion.

An examination of this equation, however, shows that if R is very large in comparison with r, so that even nr is small as compared with R, the total resistance in circuit is not very greatly increased by joining n cells in series; while the electromotive force is increased in proportion to the number of cells. Therefore, *if the external resistance is very large as compared with the internal resistance, the current is increased nearly n-fold by joining n cells in series.*

449. Batteries Connected in Multiple, or in Parallel. If, however, the n cells are joined *in multiple,* the electromotive force of the combination is the same as that of a single cell (p. 475), while the internal resistance is decreased to $\frac{1}{n}$ of that of a single cell (p. 510). That is, for n cells in multiple,

$$I = \frac{e}{\frac{r}{n} + R},$$

or

$$I = \frac{ne}{r + nR}.$$

This equation shows that *if the external resistance R is very small indeed,* so that even nR is small as compared with r, that is, if the larger part of the entire resistance in the circuit lies in the battery itself, we find that *we decrease the total resistance to nearly* $\frac{1}{n}$*th of its value, and, therefore, that we increase the current nearly n-fold by connecting the battery cells in multiple.*

Accordingly, *if the external resistance is large as compared with the internal resistance, the current will be increased by connecting cells in series. If, on the other hand, the external resistance is very small as compared with the internal resistance, the current will be increased by connecting cells in multiple.*

450. Measurement of Resistance. The Wheatstone's Bridge. The theory of divided circuits enables us most readily and accurately to measure resistance by the device known as the Wheatstone's bridge. This consists of a circuit divided into two branches through which the current likewise divides, the flow in each branch being inversely proportional to the resistance of that branch. Thus, let *ABC*, Fig. 264, be one branch and *ADC* be the other branch of this divided circuit. The fall of potential along the first branch must be just equal to the fall of potential along the second, since the ends of these branches are connected. If, then, any point such as *B* is taken on the upper branch, there will be a certain difference of potential between the points *A* and *B*. Whatever the potential of *B* may be, there will exist a point, such as *D*, somewhere on the lower branch which will be at the same potential as *B*, and therefore the difference of potential between *A* and *D* will be the same as that between *A* and *B*. If the points *B* and *D* are now connected by a conductor, *no current will flow through this wire, since both ends are at the same potential.* Accordingly, if a galvanometer, which is an instrument for detecting the flow of current, is placed in the circuit

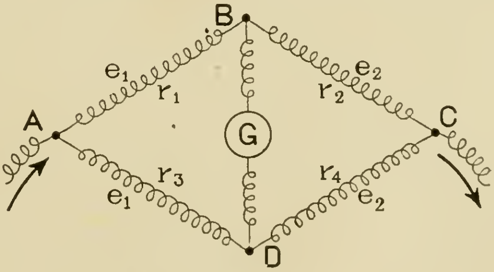

Fig. 264. The Wheatstone's bridge.

from *B* to *D*, the galvanometer will give no indication, since no current flows. If the point *D* is moved even but slightly to one side or the other, a current will flow through the galvanometer, since in this case *D* is no longer at the same potential as *B*.

Assuming now that on the lower branch the point *D* has been found which is at the same potential as point *B* on the upper branch, we see that there are four arms, or divisions to this branched circuit, namely: *AB*, *BC*, *AD*, *DC*, having resistances r_1, r_2, r_3, r_4, respectively. Moreover, the difference of potential e_1 between *A* and *B* is the same as that between *A* and *D*, since *B* and *D* are at the same potential; and the difference of potential, e_2 between *B* and *C* is the same as that between *D* and *C*, for the same reason. Furthermore, the current $\frac{e_1}{r_1}$ which flows from A to B is the same as the current $\frac{e_2}{r_2}$, which flows from *B* to *C*, since none passes through the galvanometer and there can be no accumulation of current at *B*. In other words, all the current which flows towards *B* must flow away from this point. For the same reason, the current $\frac{e_1}{r_3}$ which flows from *A* to *D* equals the current $\frac{e_2}{r_4}$ which flows from *D* to *C*. Therefore,

$$\frac{e_1}{r_1} = \frac{e_2}{r_2},$$

and

$$\frac{e_1}{r_3} = \frac{e_2}{r_4}.$$

Dividing these equations member for member, we have

$$\frac{\frac{e_1}{r_1}}{\frac{e_1}{r_3}} = \frac{\frac{e_2}{r_2}}{\frac{e_2}{r_4}},$$

or

$$\frac{r_1}{r_2} = \frac{r_3}{r_4},$$

which may be written

$$r_1 : r_2 = r_3 : r_4.$$

Therefore, when no current passes through the galvanometer, the resistances in the four arms stand in the above proportion and the bridge is said to be *balanced*. If, now, the *ratio* $r_3 : r_4$, is known, and if r_2 is a known resistance, an unknown resistance r_1 may be found from the above proportion.

451. The Slide-Wire Bridge. In practice it is convenient to make one branch of the circuit, as *ADC*, Fig. 265, of a straight, uniform wire, upon which the point of contact D may move back and forth. The resistance to be measured is introduced at r_1, and a known resistance from a resistance box or standard coil is inserted at *R*. The point *D* divides the straight wire along which it moves into two portions, l_1, and l_2, the resistances of which are directly proportional to their lengths, and these may be read at once from a scale underneath. Since in the above proportion we require simply the *ratio* $r_3 : r_4$, we may write, accordingly,

$$r_3 : r_4 = l_1 : l_2.$$

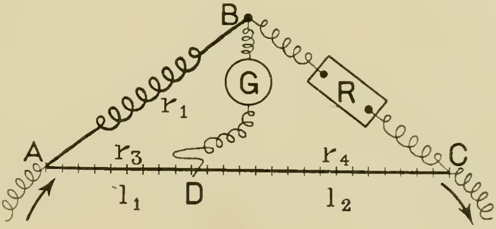

FIG. 265. The slide-wine bridge.

In operating this bridge, therefore, the wire from one terminal of the galvanometer is connected at point *B* between the resistances r_1 and *R*, while that from the other terminal slides along the straight wire AC, until no current flows through the galvanometer. This locates the point D which is at the same potential as point B, and the bridge is balanced. The ratio of the balanced bridge then becomes

$$r_1 : R = l_1 : l_2,$$

or

$$r_1 = \frac{l_1}{l_2} R.$$

Since R is a known resistance, and since the ratio $\frac{l_1}{l_2}$ is given at once by the readings of the scale, the unknown resistance r_1 is determined as a fractional part of R.

CHAPTER XXXVI.

ENERGY AND POWER.

452. The Energy of the Electric Current. In a previous paragraph (p. 496), it has been shown that a current of the intensity, I, in flowing for t seconds through a difference of potential of E units is able to perform $I\,E\,t$ ergs of work, or that

$$\text{The energy of the current} = I\,E\,t \text{ ergs.}$$

If this difference of E units of potential is maintained between the points A and B, Fig. 226, it is evident (p. 508), that between these points there must also exist a resistance, R, to which this difference of potential is

FIG. 266. A difference of potential, E, and a resistance, R, must exist between the points A and B.

proportional. Moreover, R must be of such magnitude that Ohm's law, namely:

$$I = \frac{E}{R}$$

must be obeyed by the current as the latter flow from A to B. From this expression we see that

$$E = I\,R;$$

and by substituting this value of E in the above equation, we obtain an equivalent expression for the energy of the current, that is,

$$\text{The energy of the current} = I^2\,R\,t \text{ ergs.}$$

We have, therefore, two equivalent expressions for the energy of the current, namely:

$$\text{The energy of the current} = I\,E\,t \text{ ergs,}$$

and

$$\text{The energy of the current} = I^2\,R\,t \text{ ergs,}$$

where *I, E, and R are measured in electromagnetic units.* The first of these expressions is used when the conditions of a problem are given in terms of current, electromotive force, and time. and the second is employed when these conditions are limited to current, resistance, and time.

453. The Measurement of the Energy of the Current. In the above expressions for the energy of the current, it is to be observed that the electrical quantities are measured in *C. G. S. units* and the energy is therefore, obtained in *ergs*. If, however as is usual, the current is measured in amperes, the electromotive force in volts, and the resistance in ohms, the energy will still be obtained in ergs, *provided the number of amperes, the number of volts, and the number of ohms are expressed in* *terms of electromagnetic units of current, electromotive force, and resistance respectively.* We must, therefore, remember that

1 Ampere $= 10^{-1}$ electromagnetic unit of current,

1 Volt $= 10^{8}$ electromagnetic units of electromotive force.

1 Ohm $= 10^{9}$ electromagnetic units of resistance.

Accordingly, by the first expression, $I\,E\,t$, a current of 1 ampere, in flowing for 1 second between two points which have a difference of potential of 1 volt, will develop

$$10^{-1} \times 10^{8} \times 1 = 10^{7} \text{ ergs};$$

while, by the second expression, $I^2\,R\,t$, a current of 1 ampere, in flowing for 1 second through a resistance of 1 ohm, will likewise develop

$$10^{-1} \times 10^{-1} \times 10^{9} \times 1 = 10^{7} \text{ ergs}.$$

But (p. 118),

$$10^{7} \text{ ergs} = 1 \text{ Joule},$$

from which it is evident that in either case the above currents represent 1 joule of energy, and will perform 1 joule of work.

In general, therefore, if $I =$ current in amperes, $E =$ electromotive force in volts, $R =$ resistance in ohms, $t =$ time in seconds, we have, either

The energy of the current $= I\,E\,t$ joules,

or

The energy of the current $= I^2\,R\,t$ joules.

454. The Power of the Electric Current. If the total work performed by the current is divided by the time in which this work is performed, the result is the *rate* of doing work, that is, the amount of work done in the unit of time, or the *power* (p. 119) which the current can develop. But

$$\frac{\text{Joules}}{\text{Seconds}} = \text{Watts};$$

and therefore

$$\text{The power of the current} = I\,E \text{ watts},$$

or

$$\text{The power of the current} = I^2\,R \text{ watts}.$$

455. The Current Develops Heat in Flowing Through a Conductor. Whenever a current of electricity flows through a conductor it meets within that conductor an opposition, or resistance, which impedes its flow, just as friction produces a retarding effect in mechanical motion. A certain amount of energy, therefore, must be expended in forcing the current through the conductor, and the energy thus consumed in overcoming this resistance appears in the conductor in the form of heat. Accordingly, *all conductors are heated by the passage of the electric current through them,* the amount of heat thereby produced being *proportional to the resistance encountered.*

That the heat thus developed is, indeed, a function of the resistance, may be shown by sending a current through a conductor made of alternate lengths of silver and platinum wire, in which case the current flowing through one must flow also through the other. Since in passing through this conductor, the current will encounter in the platinum a resistance about six-fold greater than in the silver, the amounts of heat developed in the two will be this same proportion, and the platinum wires will become red-hot, while those of silver will remain comparatively cool. An application of this principle is made in exploding a mine. In the charge of powder or dynamite is placed a fine platinum wire, to the ends of which are attached heavy copper wires which may be connected with a distant battery. When it is desired to have the explosion take place, the circuit from the battery is closed through this composite conductor, and the current in passing through the resistance offered by the platinum wire develops sufficient heat to cause this wire to become red hot, thereby igniting the explosive.

456. The Number of Calories of Heat Developed by the Current. In the chapter on thermodynamics we have learned (p. 323) that to develop 1 calorie of heat requires the expenditure of 427 kilogrammeters of work. That is,

$$1 \text{ calorie} = 427 \times 1000 \times 981 \times 100 \text{ ergs},$$

or

$$1 \text{ calorie} = 41{,}888{,}700{,}000 \text{ ergs}.$$

To develop 1 lesser calorie requires the 1/1000th part of this, or

$$1 \text{ lessor calorie} = 41{,}888{,}700 \text{ ergs}.$$

Since

$$10{,}000{,}000 \text{ ergs} = 1 \text{ joule},$$

we see that

$$1 \text{ lesser calorie} = 4.2 \text{ joules, nearly},$$

and, accordingly,

$$1 \text{ joule} = 0.24 \text{ lesser calories}.$$

Moreover, since (p. 518)

$$\begin{aligned} \text{The energy of the current} &= I\ E\ t \text{ joules} \\ &= I^2 R\ t \text{ joules}, \end{aligned}$$

it follows that when the energy of the electric current is expended in heating the resistance through which it flows,

The heat developed by the current $= 0.24\ I\ E\ t$ lesser calories,

or

The heat developed by the current $= 0.24\ I^2\ R\ t$ lesser calories.

The heat which is developed in a resistance by a current, and which is independent of the direction of the current, was investigated experimentally by Joule, and is known as the Joule heat.

CHAPTER XXXVII.

THE ELECTRIC LIGHT.

457. The Incandescent Electric Light. During the past thirty-five years an application has been made of the energy of the electric current, transformed into heat and light, which has wholly revolutionized our methods of artificial illumination. The idea came to Edison that if it were possible to find a conductor of electricity having a very high melting point, and at the same time offering a great electrical resistance, it might be possible to raise this conductor to a white heat by the electric current, and so obtain an artificial source of light. It is only fair to say that, even at an earlier date, others likewise had formed a similar conception, and, indeed had made lamps which could be operated in this way. The matter, however, received but little attention; and to Edison belongs the honor of developing this idea anew, and of devising the mechanism not only for the manufacture of the lamps, but also for the generation of the current. These discoveries together with many further improvements by later investigators, have brought about the modern highly-perfected system of electric lighting.

It was Edison's first idea to enclose in a glass globe a fine platinum wire, and to raise this wire to the temperature of incandescence by passing an electric current through it. While platinum possesses a fairly high melting point and the necessary high resistance, it soon developed that these essentials alone were not sufficient for producing a satisfactory filament. It was found that with continued heating platinum disintegrates, and that it also volatilizes, or sublimes, so that, even if the slender wire were not actually broken, the interior of the glass globe soon became coated with a thin film of this metal, which obstructed the passage of the light. Platinum was therefore discarded for a thread of dense carbon, which possesses in a high degree the qualities essential for an effective filament.

But carbon also was found to volatilize; and although the process was not rapid, yet in time this loss of particles from the filament so reduced the size of this conductor that the resistance increased, and thus prevented the passage of sufficient current to bring the lamp to full incandescence. Moreover, since the lamp became "blackened" by this deposit of carbon upon the interior of the globe, the already dimmed light from the filament was still further obscured by the darkened glass. Therefore, after a carbon lamp had burned about six hundred hours, it was found economical to discard it and replace it by a new one. Notwithstanding these objectionable features, carbon, for thirty years, was considered to be the material best adapted for efficient use in incandescent lighting.

Two decades ago, another most interesting and efficient incandescent lamp was devised by Professor Nernst of Göttingen. The filament of this lamp, consists of a slender rod of a material composed of various metallic oxides, not altogether unlike porcelain in general appearance, and of such properties that it is a very good conductor when hot, but almost a perfect insulator when cold. Artificial means, therefore, must be employed to raise the filament to a bright red heat before the current can pass. When the flow takes place, however, the heat developed in the high resistance of the material raises the filament to a higher temperature than may safely be attained with carbon, and the light emitted is far more intense and of a much whiter quality than that from the ordinary carbon lamp. Moreover, is is unnecessary to enclose this filament or to protect it in any way from oxygen, since it consists already of metallic oxides which cannot burn or become further oxidized by being heated in the air.

Comparatively recently, however, (1908), it has been found possible to reduce to the form of slender wires two hard, brittle metals of exceedingly high melting points, namely: tantalum and tungsten. A wire of the latter metal when used as the filament of an incandescent lamp, is heated to a temperature even higher than that of the Nernst lamp, and gives a light which is of a soft, white quality, thus making it the most desirable form of artificial illumination yet devised.

In the manufacture of incandescent lamps it has been found that the higher the temperature to which the filament is raised, the greater is the percentage of the electric energy which is transformed into light, and the smaller is the number of watts required to develop one candle power. In the following table is given approximately the temperature to which the various filaments may safely be heated, and also the number of watts of power represented by one candle power of the resulting light:

Material.	Temperature.	Watts per candle.
Platinum	1500° C.	6.0
Carbon	1850° C.	3.5
Tantalum	2000° C.	2.0
Nernst	2200° C.	1.5
Tungsten	2300° C.	1.0
Tungsten (in nitrogen)	2850° C.	0.5

From this table, it is seen that, besides the greater intensity and whiteness of the light emitted by the tungsten filament, the efficiency (p. 123) of this lamp is also very much higher than any other. Indeed, so many of the desirable qualities are found combined in tungsten, that this material bids fair to supersede all others in the incandescent lighting of the future.

It is to be observed that, with the exceptions of the Nernst "glower" and one other lamp recently invented, the filaments of all incandescent lamps are enclosed in a glass globe from which the air has been exhausted. With carbon this is obviously necessary in order to exclude the air, in which the filament would otherwise soon burn when heated. A further reason also is found for enclosing this and other filaments in an exhausted globe, namely: *to prevent a loss of heat by the convection of the air.* If the lamp were burned freely without this protection, or in a globe from which the air was not exhausted, the surrounding air in becoming heated would withdraw from the filament a portion of the energy which would otherwise be transformed into effective light radiation, and the efficiency of the lamp would be lowered.

Very recently, however, (Oct. 1913), it has been found that an incandescent lamp of much greater efficiency may be made by enclosing a tungsten filament in a glass globe which is filled with nitrogen. The pressure of this inert gas prevents the evaporation, or sublimation, of the filament which would otherwise take place even with tungsten, so that it is possible to raise the temperature of this filament to 2850° C. It is true that convection currents are set up in this gas which carry away energy from the filament; but this may be reduced to a minimum either by using a filament with large cross-section, that is, of small superficial area, or a smaller filament in the form of a closely-wound coil.

At this higher temperature, the percentage of the shorter wave lengths of light, which affect the eye, is much greater as compared with those to which the eye is insensitive. Moreover, this increase in the relative amount of energy transformed into light radiations so much more than compensates for the loss of energy by the convection currents in the nitrogen, that the efficiency of a lamp of large size constructed in this way is greater than that of any other, the power consumed being, indeed, only about 0.5 watt per candle power. The light, too, not only possesses great intensity, but also has so large a proportion of the shorter wave lengths that it has a quality which to some extent resembles daylight in character.

458. The Arc Light. If two carbon rods are connected with a high potential source of electricity and brought together, a current will flow from one to the other. If, now, the rods are slightly separated, the current is not broken, but continues to flow across the intervening space, and the ends of the carbons emit a most intense heat and a light of most dazzling brilliancy. The maintenance of the current across this space is undoubtedly due to the fact that the intense heat here developed ionizes the air and causes the carbon, if not actually to volatilize, at least to form gaseous compounds with the constituents of the air, so that the current flows from one carbon to the other across this conducting column of heated gas. It must be borne in mind that the passage of the current across this space is true conduction, and is not of the disruptive nature of a spark, because the current does not flow until the carbons have been brought into actual contact. After the separation, the space between the carbons is then filled with a column of the most intensely heated gases, and these at this high temperature are dissociated, and therefore ionized. This conducting column of heated vapour, which is nearly spherical in shape, is called the *electric arc*, and the light which is produced by the process here described is known as the *arc light*. Sir Humphrey Davy, in the early part of the last century was the first to produce the arc light,—an achievement which he accomplished by employing a battery of three thousand cells.

There are many things of remarkable interest connected with the phenomenon of the electric arc. It is to be noticed, in the first place, that the arc itself is nearly non-luminous, for at best it emits but a feeble, bluish light, while nearly all the light comes from the highly heated ends of the carbons. The latter are raised so high in temperature that they are slowly consumed by oxidation in the air, the positive carbon becoming much hotter than the negative, and wasting away about twice as rapidly. It is also to be noticed that the form which the positive carbon assumes is wholly different in appearance from that of the negative; for while the negative carbon grows into a somewhat blunt, but more or less sharply defined cone, Fig. 264, the end of the positive carbon forms into a little cavity two or three millimeters in diameter, called the *crater*. It is at the surface of this crater that the most intense heat is developed, the carbon here having, when

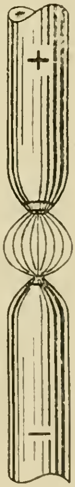

FIG. 264. The electric arc.

cold, a bright, glossy appearance, as if it had been melted under the influence of the extreme heat.

While it is true that the red-hot ends of the carbons radiate feebly, the crater of the positive carbon is so intensely heated that about ninety-five percent of the entire amount of light emitted by the arc lamp comes from this small source. Indeed, it is due to the minuteness of the source that the shadows cast by the arc light are disagreeably sharp; and as there is no gradation from the intensity of the black shadows to the full glare of the radiation from the crater, this form of light is unsuited for interior illumination, unless moderated by an opal globe, which distributes the light more uniformly. Since the carbons waste away by oxidation in the air, the mechanism of the lamp consists of a device for moving them together as fast as they are consumed, and by this means the length of the arc is maintained nearly constant. The positive carbon is usually placed above the negative, so that the light from the crater may be thrown downward to illuminate as wide an area as possible, the dark circle directly under an arc lamp being due to the shadow cast by the lower or negative carbon.

459. The Mercury Vapor Light. Still another form of the electric light, known as the mercury vapor lamp, has of late obtained some prominence, although it has not come into general use on account of the unpleasant color of the light which it emits. This lamp consists of an evacuated glass tube, often a meter or more in length, with electrodes sealed into glass at the ends, and containing a quantity of mercury. The tube is usually held in a position which is nearly horizontal, and is connected by its electrodes with a dynamo which furnishes a difference of potential of 110 volts.

When the more elevated end of the tube is tipped slightly downward, the mercury, in flowing to the lower end, establishes metallic connection between the electrodes, so that a current flows through the thread of mercury so formed. As soon as this thread breaks at any point, an arc is momentarily established. This not only volatilizes the mercury, but disengages electrons from the mercury vapor, that is, it ionizes the vapor, or fills the tube with electrons and with positively charged mercury atoms.

each of which has lost an electron. These particles, or ions, move with a high velocity because of the difference of potential between the electrodes; and in their rapid motion they collide with other atoms of mercury vapor, from some of which they likewise disengage an electron, and so the process continues.

The shock, produced by the collision between the electrons and the atoms of mercury vapor, (for the molecules of mercury vapor contains only a single atom), produces so violent a vibration of the electrons which still remain in the mercury atoms, that ether waves are set up which affect the eye with the sensation of light. Light is therefore produced by the bombardment of the atoms of the mercury vapor by the ionized parts of this vapor; and, as might be expected, the character of this light is determined by the law of the vibration of the electrons within the mercury atom. Only a comparatively few wave lengths are therefore emitted, and these are such as to produce a light of a vivid bluish-green color.

While this light is unpleasant in tint, it is, however, restful to the eye, so that it is used extensively in factories and draughting rooms, where the color is no serious objection, and where close, careful, visual work must be done.

460. Methods of Connecting Electric Lamps. On account of their high resistance, incandescent lamps cannot well be connected in series, inasmuch as it would then require a dangerously high electromotive force to send sufficient current through them. Moreover, if a filament anywhere in the series should break, the current would cease to flow, and the remaining lamps would thus be extinguished. Accordingly, it has been found advisable to connect incandescent lamps in multiple, or parallel; that is, the wires which lead from the dynamo are maintained at a constant difference of potential, and the lamps are then connected across in multiple from one main wire to the other, Fig. 265. The methods of manufacture are such as to make the filaments almost exactly the same in resistance; and since the lamps have a constant difference of potential between their terminals, the current divides equally between them, and all are uniformly brilliant.

An arc, on the other hand, requires about 40 volts for its maintenance; but its resistance, which varies with its length, depends upon many factors, such as the regulating mechanism for moving the carbons together, access of air, which causes a consumption of the carbons by oxidation, inequalities in the composition of the carbons, etc. It is obvious, then, that a number of arc lamps, each supplied with the same electromotive force, and initially otherwise identical, would soon have arcs of different lengths, and therefore of different resistances. Accordingly, if arc lamps were connected in multiple, there would soon be a very unequal distribution of current among them, the one offering the smallest resistance taking the largest current and giving the most intense illumination.

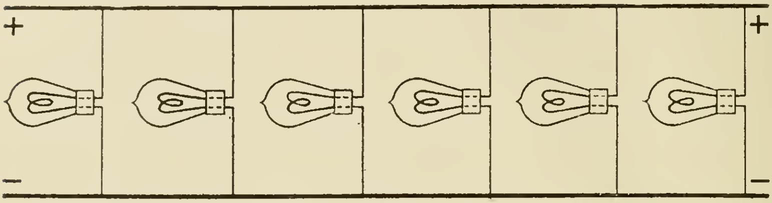

FIG. 265. Incandescent lamps are connected in multiple.

It is usual, therefore, to connect arc lamps in series, so that the same current flows from one lamp to the next throughout the entire circuit, Fig. 266. Since each lamp requires a fairly high electromotive force, or "voltage", the dynamo furnishing this current must develop a very great difference of potential if forty or fifty arc lamps are connected in series, as is the practice in street lighting. It is on this account that an arc light circuit is far more difficult and dangerous to manipulate than an incandescent circuit, for in the latter the potential, in general, does not exceed in 110 volts.

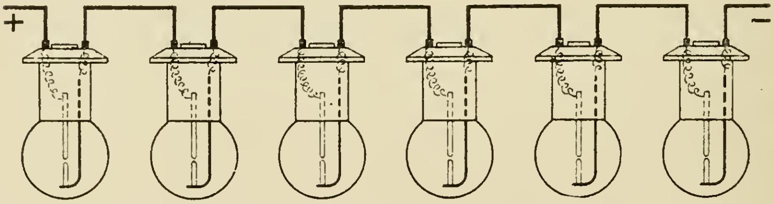

FIG. 266. Arc lamps are connected in series.

CHAPTER XXXVIII.

THERMO-ELECTRICITY.

461. The Electromotive Force of Contact. In the discussion of electromotive force it has been shown that a difference of potential is produced only by the expenditure of some form of energy. There is, however, a series of experiments which would seem to indicate that a difference of potential may be produced without the expenditure of energy in any form which we recognize as energy at the present time, and which depends upon the simple contact of two dissimilar metals.

When a plate of copper and a plate of zinc are supported by insulating handles and are simply touched together and then separated, they are found to be at different potentials. At first, this was thought to be due to a slight friction, or rubbing of one plate upon the other, but further experiment seems to render this explanation insufficient; for if the plates, insulated by varnish, are placed one upon the other, and the varnish then slightly broken, so that even a fine wire may make metallic contact between them, Fig. 267, they are found, when separated, to be at different potentials.

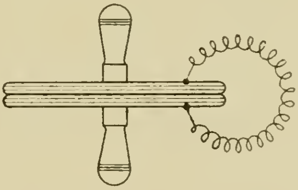

FIG. 267. The electromotive force of contact.

This production of a difference of potential simply by touching the plates together is called the electromotive force of contact, or the electrification by contact. If a plate of antimony and one of bismuth are treated in this way, the difference of potential between the two will be found to be even greater than that developed when plates of copper and zinc are used. Many physicists claim that in all cases a simple contact of two dissimilar substances produces a difference of potential, and that even the electromotive forces present in a battery are due wholly to the various contacts there existing, that is, between the zinc and the electrolyte, the copper and the electrolyte, and especially between the zinc and the copper. Indeed, this theory dates back to the time of Volta, and has had its ardent adherents ever since.

Whatever the explanation may be, the fact remains, and is not disputed by those who advocate the energy theory of electromotive force, that when plates of dissimilar substances, such as antimony and bismuth, are separated after having been brought together, the two are at different potentials.

An imperfect explanation may perhaps be given when we consider the free electrons which are present in each of the metals. From the fact that these electrons may exist at least in different numbers and possibly under different conditions in the different metals, and that the molecules of the various metals have unequal masses, we see that the conditions for the thermal equilibrium of the electrons in one metal may be entirely different from these conditions in any other metal. It is conceivable, therefore, that at the boundary between two metals the electrons from one may pass over into the other, thereby changing the potential of each. Moreover, this passage of electrons would be continued until an equilibrium condition at the boundary were established, that is, until the metals came to perfectly definite potentials.

At a higher temperature, the equilibrium condition of the electrons and the molecules in the metals may be different; and if this were the case, it would be expected that the potentials which are produced by contact would likewise be different. Indeed, these potentials are found to change as the temperature rises, and to become equal at a temperature which is different for different metals, but which is constant for any two. Above this temperature, which is known as the *temperature of inversion*, the metals assume opposite potentials, the one which was at high potential now becomes at low potential, and *vice versa.*

If we accept an electronic explanation of the contact electromotive force, of which the above paragraph is merely the crudest possible attempt, we can see that even here the production of a difference of potential is obtained at the expense of electronic energy, or, at least, by the transfer of electrons from one metal to the other. Therefore, without attempting further to explain this interesting phenomenon, we shall simply accept the fact that there does exist an electromotive force due to contact, the true nature of which, at present, is not entirely understood.

462. Thermo Electricity and the Thermo-pile. Closely allied with the production of contact electromotive force is the phenomenon of thermal electromotive force and the operation of the thermo-pile. In the year 1821, Seebeck made the discovery that if a pile is built up of alternate bars of antimony and bismuth soldered together in series, and if the first and the last bars are connected to a galvanometer, Fig. 268, no current will flow provided the junctions all remain at one temperature. If, however, the pile is so constructed that the even junctions all form one face and the odd junctions the other, and one face is heated or cooled, a current will flow one way through the galvanometer when this face is heated, and the other way when this face is cooled. It will be noticed that the direction of the current thus generated is always *from the bismuth to the antimony across the heated junction.*

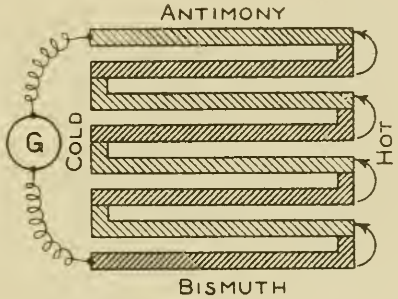

FIG. 268. The construction of the thermo-pile.

Such an instrument is called a thermo-pile, and is exceedingly sensitive in detecting minute changes of temperature between the faces. So delicate is this instrument as a heat detector, that a deflection of the galvanometer may be observed when the temperatures of the two faces do not differ by more than one-millionth of a degree C. The difference of potential developed by a thermo-pile is very small as compared with that produced by a battery, and hence it is not infrequent to connect as many as a hundred pairs of antimony and bismuth bars in series in a single thermo-pile, in order that the resulting difference of potential may be greater. This vastly increases the usefulness of the thermo-pile as a source of electricity, because the electromotive force is thereby multiplied, and the internal resistance of the pile even then is very low, inasmuch as the circuit consists entirely of metal.

It is of interest to observe that the electromotive force of the thermo-pile depends not only upon the *difference* in temperature between the two series of junctions, but also upon the *actual* temperatures of these injunctions. Whatever may be the thermal electromotive force for ordinary temperatures, this electromotive force is found, in general to rise to a maximum value and then to decrease, as the temperature of one series of junctions is continuously raised, until a temperature is reached at which the electromotive force becomes zero. Above this temperature, which is the *temperature of inversion* for these metals, the electromotive force again increases, but with opposite sign; that is, the thermo-pile reverses its polarity, and gives rise to a current flowing in the opposite direction.

463. The Energy of the Thermo-current. At moderate temperatures, although the current flows from the bismuth to the antimony across the heated junction, the bismuth when tested is found to be at low potential, while the antimony is at high potential. There is, therefore, at the point of contact, a sharp change from the potential of the bismuth to that of the antimony. Accordingly, if we represent the potential of the bismuth by the lower line, Fig. 269, and the potential of the antimony by the upper line, the difference of potential at the heated junction is indicated by the perpendicular distance between the two.

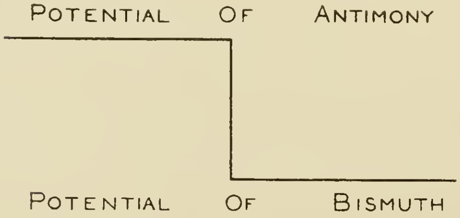

Fig. 269. The difference of potential between bismuth and antimony at the heated junction.

The application of energy in the form of heat at this junction causes a movement of electrons from the high potential antimony to the low potential bismuth; that is, it causes a flow of the current from the low potential bismuth to the high potential antimony. We see that this is consistent with our knowledge of the flow of the electrons, since we know that these particles flow *of themselves* from a region of low potential to a region of high potential, and perform work in so doing. However, just as it is possible by the expenditure of mechanical energy to carry electrons from a high to a low potential, so also may we conceive that, by the expenditure of thermal energy, electrons may be made to flow from a region of high potential to one of lower potential. This, indeed, is the case here; and while electrons thus flow from the high potential antimony to the low potential bismuth, *this flow is inevitably and invariably associated with the absorption of heat energy at this junction.*

464. The Peltier Effect. The closing remark of the previous paragraph shows that current cannot flow from the bismuth to the antimony, that is, that electrons cannot flow from the high potential antimony to the low potential bismuth, without the necessary absorption of heat at this junction. Not long after Seebeck's discovery, it was shown by Peltier that, inasmuch as the heat which is taken up at this junction causes a current to flow from the bismuth to the antimony, it must follow that *if a current from an outside source is sent from the bismuth to the antimony, heat must be absorbed at this junction, and if no heat is supplied, this junction must grow cold.* Referring again to the last diagram, we see that this result, which is known as the Peltier effect, must be so; for, if a current is sent across this junction from the bismuth to the antimony, we force the electrons to pass from the high potential antimony to the low potential bismuth, and this can be effected only by the expenditure of a certain amount of energy, or by the absorption of heat. Therefore, if no heat is supplied at this junction, the energy which is thus required to force the electrons to pass from the high potential antimony to the low potential bismuth, must be taken from the heat which is in the metals themselves, and the junction accordingly grows cold.

If the other end of the bar of antimony is joined to another bar of bismuth, Fig. 270, the potential falls as the current passes to this second bar; and here, as the electrons pass from the low potential bismuth to the high potential antimony, they not only require no energy to make them flow, but they flow in this direction *of themselves, and, in so doing, give out energy, so that this junction grows warm.* Accordingly, *if a current from an outside source is sent through a thermopile, one face will grow cold and the other warm,* the face which grows cold being the junctions where the current passes from the bismuth to the antimony, and the face which grows warm being the junctions where the current flows from the antimony to the bismuth.

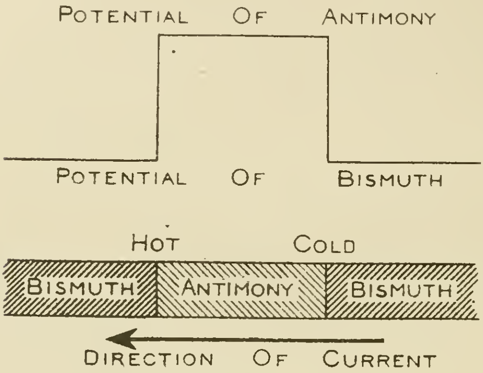

Fig. 270. The Peltier effect.

465. Demonstration of the Peltier Effect. The demonstration of the cooling and the warming phenomena due to the Peltier effect may be easily accomplished: a bar of bismuth is soldered to a bar of antimony, and the other end of the bar of antimony is soldiered to a second bar of bismuth, Fig. 271. If a current from an outside source is sent through this compound bar, the first junction will become cooled, while the second junction will grow warm, due to the causes which have just been described. If an air thermometer is placed around each of the two junctions, the air in the first will be cooled by the cold junction and so will contract, while the air in the second will be heated due to the hot junction, and so will expand.

But the current in passing through the resistance which this conductor offers will heat the conductor, and so will cause the air in each thermometer to expand. This Joule heat, or $I^2 R$ heat, which is thus developed by the current due to the resistance of the conductor, is the same in each air thermometer, and therefore if the lower parts of the tubes of the air thermometers are connected together, the tendency of the air in one bulb to expand and push the index one way will be counterbalanced by an equal tendency of the air in the other bulb to expand and force the index the other way. The movement of the index, therefore, will be wholly uninfluenced by the Joule, or $I^2 R$, heat which is developed, and will be affected only by the heat or cold due to the Peltier effect. Accordingly, if the current is sent through this compound bar from right to left, the right hand junction will grow cold and the other will grow warm due to the Peltier effect, and this will be indicated by the contraction of the air (the rising of the index) in the right hand air thermometer, and the expansion of the air (the falling of the index) on the left hand side.

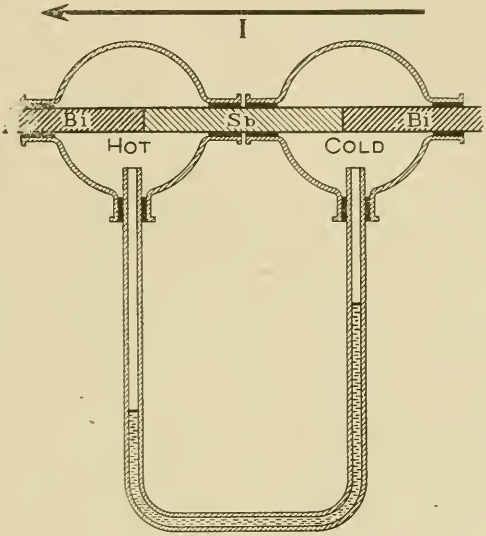

FIG. 271. Demonstration of the Peltier effect.

If the current in this apparatus is reversed, the junction which was formerly heated now becomes cooled, and that which before was cooled now becomes heated. In other words, the Peltier effect is reversed when the direction of the current is reversed, that is, *the Peltier effect is a phenomenon which depends wholly upon the direction of the current.* In this respect the heating or cooling of a junction due to the Peltier effect is different from the heat developed by a current passing through a resistance; for while the former depends upon the direction of the current, the latter is wholly independent of the direction in which the current flows.

Another and possibly a simpler method of showing the peltier effect is to send the current from a battery through an ordinary thermo-pile and a galvanometer in series, and to observe by the deflection of the galvanometer the direction in which the current flows. After this flow has continued for a short time, the battery may be removed, and the thermo-pile connected directly with the galvanometer. In proof of the Peltier phenomenon it will then be observed that *a current flows from the thermo-pile as a source, thus showing that the two faces of the pile had been brought to different temperatures by the passage of the current through it. Moreover, this current flows in direction opposite to that which was sent through the pile,* thereby proving that the heating and the cooling of the junctions produced by the current was opposite to that which would have to be applied to the pile to *generate* a current in the same direction. That is, the junctions which would have to be heated to generate a current in a certain direction had been cooled by sending a current in this direction through the pile. In other words, the flow of current across these junctions is inevitably associated with the absorption of heat at these junctions.

CHAPTER XXXIX.

ELECTROMAGNETISM

466. Oersted's Discovery. The year 1819 is a memorable one in the history of electrical science. In that year, Oersted, Professor of Physics at Copenhagen, made a discovery so fundamental in its importance and so fruitful in its results that it opened an entirely new field for electrical investigation, and since that date it has turned the tide of thought and research along electrical lines into channels hitherto unknown.

Oersted's remarkable epoch-making discovery was the simple fact that a wire carrying an electric current is surrounded by a magnetic field. From this it follows that a magnetic needle placed in the presence of such a wire is acted upon by a force which deflects the needle. Oersted found that the direction in which the needle turns depends upon the direction of the current and the position of the wire with reference to the needle. These experiments, which were repeated by many learned men of that period, proved to be of unusual interest to all, and to no one more so than to the young French scientist, Ampère. This great physicist not only confirmed the results heretofore announced, but discovered many truths regarding the magnetic field about an electric current. Moreover, his investigations enabled him to formulate the laws according to which these magnetic fields react upon one another and upon other magnetic fields, which are known as Ampère's laws at the present time.

467. Ampere's Rule for Determining the Direction of the Current. It was Ampere who stated the rule which still bears his name showing the direction of movement of the magnetic needle with relation to the direction and position of the wire carrying the current. In determining the mutual action between a current and the magnetic needle, the conductor carrying the current is in all cases to be held in the magnetic meridian, and therefore parallel to the needle. Ampère's rule is then as follows: *If you swim in the current and with the current, facing the needle, the north pole of the needle will be deflected towards your left.* Thus, if the wire is above the needle, and the current is flowing from south to north, the little swimmer will swim with his face downward, and the north pole of the needle will accordingly be deflected west. If the wire is so turned as to cause the current to flow south, and the wire is placed under the needle, the little swimmer will swim south with his face upward (facing the needle), and the north pole of the needle will again be deflected to the west.

Accordingly, if the wire is bent in the form of a loop with the needle at the center, and this loop is placed with its plane in the magnetic meridian, the current will flow in one direction above the needle and in the opposite direction below it, in which case both parts of the current will produce a deflection in the same direction, and the needle will then be deflected with a two-fold force. In a similar way it is seen that if the wire is carried above the needle, then back below it, then bent so as again to pass above it, Fig. 272, and so on, or in other words, if a coil of wire is made and the

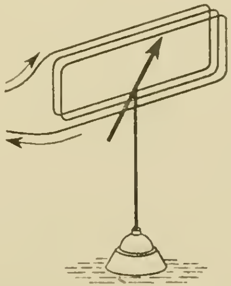

FIG. 272. The deflection of a magnetic needle by an electric current.

needle placed at its center, the effect which one turn of wire exerts upon the needle is multiplied by the number of turns forming the coil, and the action of the current upon the needle may thus be increased almost indefinitely.

468. The Direction of the Magnetic Field About a Current. A careful application of Ampère's rule enables us to determine the relation which exists between the direction of the current and the direction of the lines of force in the magnetic field about the current. If the black spot in the first diagram, Fig. 273, represents the cross-section of the conductor carrying a current flowing towards the plane of the paper, away from the observer, the direction of the arrows indicates the direction in which the north pole of the needle will turn when the needle is in the different positions about the wire. In a similar way the black spot in the second diagram, Fig. 273, indicates the cross-section of a conductor carrying a current which flows from the plane of the paper towards the observer, and the direction of the arrows indicates in each case the direction in which the north pole of the magnetic needle will move. At positions intermediate between those here specified, the needle assumes intermediate directions; and if the needle is carried continuously around the wire, the north pole will be found to describe a complete circle, the plane of the circle being at right angles to the direction of the wire, Figs. 274 and 275.

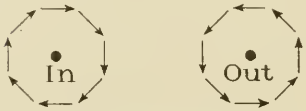

FIG. 273. Positions assumed by the needle with reference to a current.

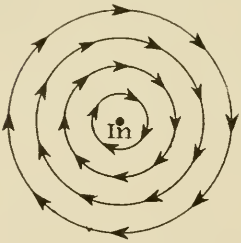

FIG. 274. Lines of force about a current flowing in.

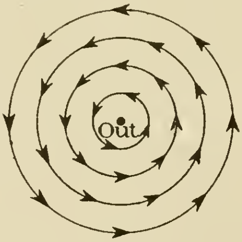

FIG. 275. Lines of force about a current flowing out.

The fact that the needle is thus acted upon by a force indicates that a conductor carrying a current is surrounded by a magnetic field. Moreover, since the direction of the lines of force in a magnetic field is indicated at each point by the direction in which the north pole of a magnetic needle moves when placed at that point, we see that the lines of force about a wire carrying a current are circles, the plane of the circles being everywhere perpendicular to the direction of the conductor carrying the current.

If a hole is bored in a piece of glass and a wire carrying a current is brought through this hole, iron filings sifted upon the glass will arrange themselves in the form of concentric circles about the wire, showing that the field of force about a current is a circular one, Fig. 276. Moreover, if a wire through which a current is flowing is dipped into a mass of iron filings, the filings will cling to the wire, not in the form of tufts as they cling to a magnet, but in a comparatively smooth, compact mass, thus showing also that the lines of force do not radiate from a wire, as they do from the ends of a magnet, but that they exist like so many rings in the form of circles about the wire and concentric with it.

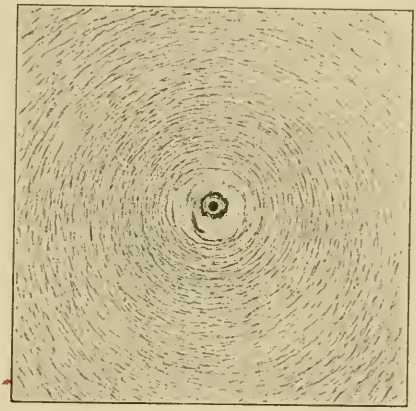

FIG. 276. The circular magnetic field about a current, as mapped by iron filings.

After examining diagrams similar to Figs. 274 and 275, Maxwell formulated the following rule which enables us at once to determine the direction of the lines of force about a current when the direction of the current is known: If you were driving a screw in the direction in which the current is flowing, then the lines of force exist about the wire in the direction in which you would have to turn the screw. Another rule for the same purpose is found very serviceable by many: If you look along the conductor in the direction in which the current is flowing, the lines of force will then extend about the conductor in the direction of the motion of the hands of a watch. Still another convenient and simple means of determining the direction of the magnetic field about a current is to grasp the conductor with the right hand, the thumb extended in the direction in which the current is flowing, in which case the fingers are then wrapped about the conductor in the direction of the lines of force.

469. The intensity of the Magnetic Field About a Current. From what has been said regarding the magnetic field about an electric current, it is evident that the effect of the current is produced not only within the conductor, whether this is solid or liquid, but in the entire space around the conductor; for the entire region in the neighborhood of a conductor carrying a current is filled with lines of force which find their origin in the current. Indeed, these two phenomena are invariably associated; so that the existence of a current in a conductor implies a field of force about that conductor, and, conversely, the presence of a field of force about a conductor is indicative of the flow of current within. Therefore, since these lines of force about a current are always intimately connected with the current itself, the strength of the current may be measured in terms of the action of this magnetic field upon a magnetic needle of known strength.

It has been found by experiment that when all parts of a current are brought to the same distance from a magnetic needle, the force with which the magnetic field due to the current acts upon the magnetic needle is proportional to the strength of the current, to the length of the current, to the strength of the pole of the needle, and is inversely proportional to the square of the distance between the needle and the current. Accordingly, if I denotes the strength of the current, L the length of the current, that is, the length of the conductor carrying the current, m the strength of pole of the needle, r the distance between the current and the needle, and F the force with which this current acts on the needle, then

$$F : \frac{I\,L\,m}{r^2}$$

or

$$F = k\frac{I\,L\,m}{r^2},$$

where k is the factor of proportionality.

470. The Electromagnetic Unit of Current. In this expression, which shows the relations between the current which flows and the force with which this current acts upon a magnetic needle, the units of all the quantities involved are known except that of the current, while the value of the constant, k, remains likewise unknown. However, if we agree to call *the current unity, when the length, the force, the magnetic pole, and the distance of the current from the pole are each equal to unity,* then

$$k = 1,$$

and the above expression becomes

$$F = \frac{I\,L\,m}{r^2}.$$

In order to give to k the value of unity in this expression, we have thereby limited the value of the unit current to the above conditions. Therefore, *if all parts of a current are brought to the same distance from a magnetic pole, the electromagnetic unit of current is defined as a current of such intensity that unit length of this current, acting at unit distance from the unit magnetic pole, will exert unit force upon that pole.* All parts of the current may be brought to the same distance from the needle by bending the conductor into the arc of a circle and placing the needle at the center. Practically, therefore, *the electromagnetic unit of current is defined as a current of such intensity that unit length of this current, bent into the arc of a circle of unit radius, will exert unit force upon a unit magnetic pole placed at the center of this arc.* In other words, *the electromagnetic unit of current is defined as a current of such intensity that a length of 1 centimeter of this current, bent into the arc of a circle of 1 centimeter radius, will* *exert a force of 1 dyne upon a unit magnetic pole placed at the center of this arc.*

The unit of current thus defined is known as the *electromagnetic unit of current*, because its determination depends upon the magnetic relations of the electric current. This is the electromagnetic unit of current to which frequent reference has been made in the preceding pages (pp. 482, 495). Moreover, because the determination of the electromagnetic unit of current is so fundamental, the entire system of units which has to deal with the flow of electricity, is known as the electromagnetic system.

As has been pointed out (p. 482), this unit at the time of its adoption was thought to be too large for ordinary use and hence 1/10 of this electromagnetic unit was taken as a practical or working unit, and was called the ampere. Therefore, the ampere is *defined as 1/10, or 10^{-1}, electromagnetic unit of current.*

471. The Sequence of the Electrical Units. A review of the electrical units will show that in many instances one unit cannot be defined without involving also the definition of another unit. Accordingly, in defining these units, a certain sequence must be observed, for it is evident that one quantity cannot be defined unless all other quantities which enter into its definition have themselves been defined.

From an examination of these electrical units, it becomes evident that the electromagnetic unit of current is the fundamental unit, inasmuch as it enters into the definition of each of the other units. After this has been defined, as above, in terms of units, it may then be used to determine the electromagnetic unit of electromotive force or difference of potential, inasmuch as this unit (p. 497) is expressed in terms of the electromagnetic unit of current, the unit of time, and the unit of work or energy, all of which are now known. When the electromagnetic units of current and electromotive force are thus defined, then, and not until then, may the electromagnetic unit of resistance be defined (p. 503) through the relations that hold true in Ohm's law. Moreover, the electromagnetic unit of quantity of electricity may now be defined (p. 494), since this unit depends upon the definitions of the electromagnetic unit of current and the unit of time. Finally, when the electromagnetic units of quantity of electricity and difference of potential have been defined, it is then possible to define the electromagnetic unit of capacity (p. 498), which is expressed in terms of the units of quantity and potential.

472. Table of the Principal Electrical Units. The following table, which gives the definitions of the electromagnetic units, and which shows the relations between these units and the working or practical units, may be found convenient for future reference.

The electromagnetic unit of current is defined as a current of such intensity that a length of 1 centimeter of this current, bent into the arc of a circle of 1 centimeter radius, will exert a force of 1 dyne upon a unit magnetic pole placed at the center of this arc (p. 541).

The ampere, which is the practical or working unit of current, is defined as 1/10 or 10^{-1} electromagnetic unit of current (p. 542).

The electromagnetic unit of difference of potential or electromotive force is defined as a difference of potential of such magnitude that it will enable 1 electromagnetic unit of current in 1 second of time to perform 1 erg of work (p. 497).

The electromagnetic unit of difference of potential is $\frac{1}{30,000,000,000}$, or 3×10^{-10}, electrostatic unit of difference of potential (p. 496).

The volt, which is the practical or working unit of difference of potential or electromotive force is defined as a difference of potential equal to 100,000,000, or 10^8 electromagnetic units of difference of potential (p. 497). The volt is thus $\frac{1}{300}$ or 3×10^{-3}, electrostatic unit of difference of potential.

The electromagnetic unit of resistance is defined as a resistance of such magnitude that 1 electromagnetic unit of difference *of potential applied to its terminals will cause 1 electromagnetic unit of current to flow through it* (p. 503).

The ohm, which is the practical or working unit of resistance is defined as the resistance through which an electromotive force of 1 volt will cause a current of 1 ampere to flow. It is equivalent to 1,000,000,000, or 10^9, electromagnetic units of resistance (p. 504). This resistance is *represented* for practical purposes by the *international ohm,* which is defined as "the resistance offered to an unvarying electric current by a column of mercury at the temperature of melting ice, 14.4521 grams in mass, of a constant cross-sectional area, and of a length of 106.3 centimeters" (p. 505).

The electromagnetic unit of quantity of electricity is defined as the quantity of electricity which passes through a circuit when 1 electromagnetic unit of current flows for 1 second (p. 494).

The electromagnetic unit of quantity of electricity is 30,000,000,000, or 3×10^{10} electrostatic units of quantity (p. 494.)

The coulomb, which is the practical or working unit of quantity is defined as 1/10 of the electromagnetic unit of quantity. The coulomb therefore is the quantity of electricity which passes through a circuit when 1 ampere flows for 1 second (p. 495). It is equal to 3,000,000,000, or 3×10^9 electrostatic units of quantity (p. 482).

The electromagnetic unit of capacity is defined as the capacity of a conductor or a condenser which 1 electromagnetic unit quantity of electricity will raise to 1 electromagnetic unit of potential (p. 498).

The electromagnetic unit of capacity is 9×10^{20} electrostatic units of capacity.

The farad is defined as the capacity of a conductor or a condenser which 1 coulomb of electricity will raise to the potential of 1 volt. The farad is therefore 10^{-9} electromagnetic units of capacity (p. 499).

The microfarad, which is the practical or working unit of capacity, is $\frac{1}{1,000,000}$, or 10^{-6}, farad. It is therefore 10^{-15} electromagnetic units of capacity. The microfarad is thus seen to be equal to 9×10^5 electrostatic units of capacity.

473. A Circular Current—The Equivalent Magnetic Shell. A most important conclusion is drawn from the fact that a magnetic field of force exists around a conductor through which a current is flowing, and that the lines of force about this conductor are circles, the planes of these circles being everywhere perpendicular to the direction of the current. If a wire carrying a current is bent into the form of a loop, we shall find, by applying Maxwell's rule for determining the direction of the field, that the lines of force due to the current all flow together through the interior of the loop, passing in at one face and out at the other, Fig. 277, and that they then bend around the outside of the wire and so form closed curves about the conductor. These lines of force, all passing in at one face of the loop and out at the other, produce a field of force about the loop precisely similar to that about a magnetic shell (p. 432).

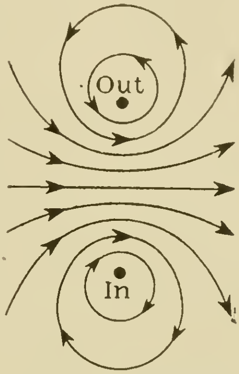

Fig. 277. The field of force about a circular current—the equivalent magnetic shell.

Accordingly, a circular current is equivalent to a magnetic shell and may everywhere replace the shell, provided the conductor through which the current is flowing coincides in position with the boundaries of the shell, and the strength of the current is such that the intensity of the magnetic field at any point due to the current is the same as that which would be produced by the shell. From this it is seen that the equivalent magnetic shell has poles, the face of the circle from which the lines of force emerge being the north pole, and that into which they pass being the south pole. It follows, therefore, that a circular current, if free to move, will set itself in the earth's magnetic field so that the north pole of the equivalent magnetic shell thus formed will point north, and the south pole will point south.

In the purely imaginary magnetic shell, where this is composed of large numbers of little magnets, we may follow the lines of force only from the north pole around the edges of the shell to the south pole, while in the equivalent magnetic shell produced by a circular current, we may follow these lines continuously as they pass in closed curves around the conductor. The north pole of a magnetic needle placed in the center of an equivalent magnetic shell will be urged in the direction of the lines of force of the shell, while the needle itself will tend to assume a position parallel to these lines of force, that is, perpendicular to the plane of the circle through which the current is flowing.

474. The Galvanometer. The presence of a magnetic field around a conductor, and the strict proportionality between the strength of this field and the current which flows enable us to measure the intensity of the current in terms of the force which this magnetic field exerts upon a magnetic needle. Any device thus for measuring a current in terms of its action upon a magnetic needle is called a galvanometer, and an instrument constructed upon this principle forms one of the two types of galvanometers in common use for measuring and detecting currents.

There are two general classes of needle galvanometers: First, those used for absolute measurements, which enable currents to be expressed in terms of the deflection of the needle and certain constants connected with the instrument itself, and second, those in which every precaution is taken to secure as large a deflection as possible from very feeble currents, and which may be calibrated by observing the deflection produced by a known current. Galvanometers of the first class are useful for determining very accurately comparatively heavy currents, which are surrounded by relatively strong magnetic fields, while those of the latter class form very delicate instruments capable of detecting currents almost indefinitely small. Formerly, galvanometers of the first class formed practically the only known device for the accurate measurement of relatively strong currents, and instruments of enormous size were constructed for this purpose. In recent years, however, other methods have been employed, and so the galvanometer as a primary instrument has now become obsolete. On the other hand, galvanometers of the second class are still extensively used, and occupy so important a place in the detection and measurement of feeble currents that the various principles upon which their action depends will be explained in detail.

475. Sensitive Galvanometers. When we have to do with exceedingly weak currents, as is often the case, the great object then becomes to devise a means of detecting even the existence of these currents; and therefore every precaution is taken to cause the magnetic fields about these feeble currents to be as effective as possible in their action upon the magnetic needle. Many points have to be taken into account if the galvanometer is to be made *sensitive,* so that even a very feeble current will produce a measurable deflection of the needle.

There are two general types of sensitive galvanometers in use, the one, a form of the needle galvanometer, originally devised by Sir William Thomson, the late Lord Kelvin, and the other, the D'Arsonval, or suspended-coil type. Only the first of these will be described here, while the D'Arsonval galvanometer will be discussed later (p. 559), after the principle upon which it depends has been more fully treated.

The first and possibly the most obvious point to consider in making a sensitive galvanometer of the Kelvin type is *to wind about the needle as many turns of wire as possible.* It has been seen (p. 538) that the effect upon the needle is proportional to the length of the conductor, or that the effect of one turn of wire is multiplied by the number of turns employed; and hence in the construction of a sensitive galvanometer often many thousand turns of wire are wound about the needle, in order that the resulting field due to the current may be as intense as possible.

Another point to be observed is *to use wire of as good conductivity and of as large diameter as is consistent,* in order that the resistance of the galvanometer may not be too great. This is necessary since the resistance of the galvanometer itself is part of the resistance through which the current has to flow; and in order to obtain the maximum current, the resistance must be kept as low as possible. We have also seen (p. 540) that the force exerted by the current upon the needle varies inversely as the square of the distance between the wire and the needle, and therefore *it is of the greatest importance to have the wire which carries the current brought very close to the needle upon which it acts.* Therefore, in a sensitive galvanometer there is left barely sufficient space for the needle to swing, the wire being wound as near the needle as possible without touching it.

The means of supporting the needle is another matter to which great attention has been given in the construction of a sensitive galvanometer. Formerly, the needle was mounted upon a pivot, but it was soon found that the friction of the latter introduced a force which tended to prevent the needle from turning when a feeble current passed. The needle was then suspended by a single fiber of silk, as this is unwound from the cocoon; but a new difficulty arose from the fact that the silk does not have perfect power of restitution (p. 114), that is, it does not come back to its original position after it has been twisted through an angle, or at best it does so but slowly. It has been found, however, that quartz may be melted and drawn out into a thread finer than a spider's web, and this is not only perfectly elastic but relatively stronger than steel. Therefore *for an exceedingly delicate galvanometer the needle is suspended by a quartz fiber.*

Another precaution which is taken in the construction of this class of galvanometers is *the rejection of the pointer,* with which the needle was formerly provided and by which the deflection was read, and the substitution in its place of *a small mirror upon which a beam of light may be thrown and reflected to a scale.* The old form of pointer, by its mass and inertia, prevented the needle from responding to the action of a small current; but the beam of light which is thrown upon the mirror of the modern galvanometer may be made of indefinite length, and as it has no mass, it has also no inertia. Possibly no one single improvement in the construction of sensitive galvanometers is of greater value than this.

Another advance that has been made in securing the greatest effect from the current is in *the use of the astatic needle* in place of the simple needle. The couple due to the current tending to turn the needle out of the magnetic meridian is always opposed by a couple due to the earth's field which tends to turn the needle again into the magnetic meridian. The astatic needle is a device whereby the full effect of the current is retained, while the effect of the earth's field is very much diminished. This needle consists of two magnetic needles, as nearly alike as possible, mounted rigidly upon a connecting rod, with their similar poles pointing in opposite directions, Fig. 278. If the two needles were exactly alike, this astatic system would remain in equilibrium in whatever position it might be placed. In practice, however, one needle is always slightly stronger than the other, and hence its *resultant* pole, which alone is affected by the earth's field, is the difference in the strengths of the poles of the two needles thus mounted together.

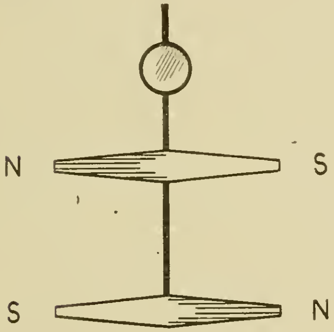

FIG. 278. The astatic needle.

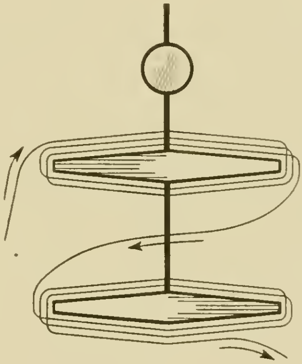

FIG. 279. The flow of current about the astatic needle.

The astatic needle, therefore, tends to set itself north and south with a very feeble force, which, indeed, is only the difference between the effects of the earth's field upon the two needles, and this difference may be made exceedingly small by making the needles very nearly alike. However, by winding the wire one way around one needle and the other way around the other needle, Fig. 279, *the full effect of the current, is obtained upon each needle separately;* and since the poles of these needles are in opposite directions, the forces exerted by the current upon both needles are in the same direction. *The effect of the current upon the astatic needle is therefore summational,* that is, the full effect of the current upon one needle is added to the full effect of the current upon the other needle. The *effect* *of the earth's field upon the astatic needle, however, is only differential;* for the effect of the earth's field upon the stronger needle is *almost neutralized* by its opposing effect upon the weaker needle.

The entire suspended system of an astatic needle galvanometer must be of exceedingly small mass in order to have the smallest possible moment of inertia (p. 142). Indeed, in a good galvanometer of this type it is frequently the case that the entire moving system, consisting of the two magnetic needles, the connecting bar, and the mirror with its mountings, has a mass which does not exceed a single milligram.

Another effective point in the construction of a delicate galvanometer is *the use of a controlling, or directing, magnet.* This is a weak bar magnet, often placed over the galvanometer itself, whereby the action of the earth's field upon the astatic system may be still further diminished. It is found practically impossible to make an astatic needle so nearly balanced that the earth's field will have the desired feeble effect. The equivalent of this, however, is accomplished by bringing the controlling magnet to such a point *that the opposing field of this magnet may almost, but not quite, neutralize the earth's field;* that is, the resultant field at the astatic needle, which is the difference between the earth's field and that of the magnet, will be almost zero. Indeed, if the controlling magnet, is approached still nearer, so that the influence of its field upon the needle is slightly stronger than that of the earth's field, the resulting field at the needle may be made as weak as desired, and yet the field of force given another direction, which is often convenient. It is when used in this way that the controlling magnet receives also the name of the directing magnet.

476. Comparison Between the Galvanometer and the Voltameter. It has already been pointed out (p. 488) that the deposition of a metal, such as copper or silver, upon the negative electrode of an electrolytic cell is an excellent means of determining the strength of the current. In this measurement it is assumed that the unit current, as expressed in C. G. S. electromagnetic units (p. 541), has been previously determined, and that the electrochemical equivalent (p. 487) of the metal to be deposited has been carefully ascertained. Either the voltameter (p. 488) or the galvanometer may be used indifferently for measuring current, although there are certain inherent differences in these two methods which must constantly be borne in mind. *The galvanometer measures the intensity of the current at any instant,* and as the current varies in intensity the deflection of the needle varies also. The galvanometer, therefore, *not only indicates the strength of the current at any instant of time, but shows to what extent the current is varying,* if this is not constant.

The voltameter, on the other hand, measures quantity of electricity, and indicates, by the amount of metal deposited, *the total quantity of electricity that has passed for the entire time during which the current has been flowing.* If the current fluctuates in intensity, a greater or a less amount of metal is deposited as the current becomes stronger or weaker, and thus *only the average of the current for the entire time may be obtained.* The voltameter, therefore, produces only an integral effect, and gives no indication of the fluctuations of the current or the magnitude of these fluctuations, except as these would modify the average of the whole. The difference between these two instruments may be expressed by saying that *the voltameter measures the quantity of electricity that has been flowing for the entire time, and determines the average, or equivalent, current, while the galvanometer measures the strength of the current at every instant, and indicates any changes or fluctuations.* In terms of our ordinary units, *the voltameter measures total coulombs, while the galvanometer measures amperes.*

CHAPTER XL.

ELECTRODYNAMICS.

477. The Mutual Action of Currents. Since every electric current is surrounded by a field of force, it is evident not only that a current must act upon a magnetic needle placed near it, but also that two neighboring electric currents must react upon each other. The nature of the mutual action which two currents are able thus to exert becomes at once evident if we consider the fields of force which exist about these currents. These fields will produce certain resultant effects depending upon the direction of the lines of force, and therefore upon the direction of the currents in the conductors. In considering these effects there are three cases to be considered.

478. When the Currents Flow Parallel and in the Same Direction. Since the currents flow parallel and in the same direction, the circular magnetic field about one current has the same direction as the similar field about the other. If we map out by iron filings the resultant magnetic field which these two currents produce, Fig. 280, we shall find that the lines of force nearest the conductors are circles, but that those farther away, become somewhat eccentric and elongated, and that finally a line of force, after passing around one conductor, will also encircle the other. Beyond this point, it is found that all the lines of force extend around the two conductors in common, thereby binding them together; and since we attribute to lines of force the

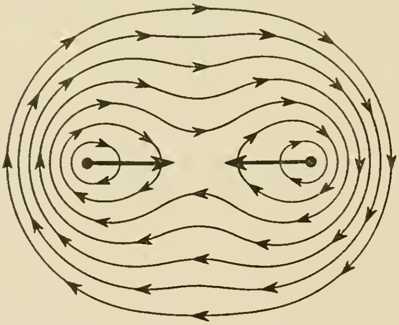

FIG. 280. Lines of force about two currents flowing parallel and in the same direction.

properties of tightly stretched rubber cords, these lines, as they contract, draw together the conductors about which they are wrapped. Therefore, *two currents which flow parallel and in the same direction attract each other.* Moreover, it is seen that this attraction is entirely magnetic, and is due wholly to the contractile tendency of the lines of force which are common to the two currents.

479. When Two Currents Flow Parallel and in Opposite Directions. Inasmuch as the currents are now flowing parallel but in opposite directions, the circular magnetic field about one is in a direction opposite to that which surrounds the other. In this instance, also, if we map out by iron filings, the resultant magnetic field produced by these two separate fields, we shall find that the lines of force due to both currents pass through the space between the two currents, Fig. 281. Moreover, the lines of force which are thus crowded together in this limited region, are parallel and in the same direction; and, as they can never intersect, we know by the properties of such lines of force (p. 424), that they repel one another laterally. We shall find, accordingly, that in the immediate vicinity of the conductors the lines of force are circles; but at a greater distance, these curves about each conductor become eccentric, and are elongated on the farther side. Each line of force, therefore is crowded by the other lines of force that pass through this narrow space; and as lines of force thus crowd or repel one another laterally, they must accordingly drag with them the conductors about which they are wrapped, so that *the conductors themselves are forced apart.* Thus it is seen that *two currents flowing parallel and in opposite directions repel each other.* Moreover, the cause of this repulsion is to be found not in the action of one *current* upon the other, *but in the mutual reaction of the fields of force by which these currents are surrounded.*

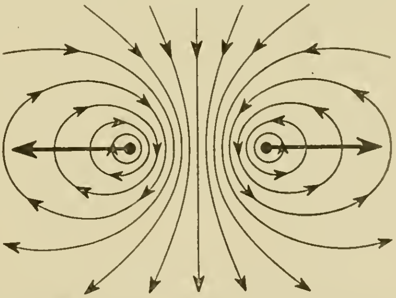

FIG. 281. Lines of force about two currents flowing parallel and in opposite directions.

480. When Two Currents Make an Angle with Each Other. In this case we shall find, by a reasoning similar to that just given, that if both currents flow towards an angle or both flow away from an angle, Fig. 282, the magnetic fields will be such that the two currents will attract each other and tend to make the angle smaller; but if one current flows towards an angle and the other away from this angle, the magnetic fields will so react that these currents will repel each other and tend to make the angle larger. Both of these results may be expressed by saying that *when currents make an angle with each other, they tend so to set themselves that they may flow parallel and in the same direction.*

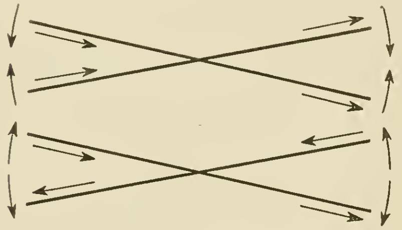

FIG. 282. Mutual reactions of angular currents.

In all the cases here cited, the results are those which were investigated experimentally by Ampere, and which are embodied in the three laws which are still known as Ampere's laws, as follows:

1. *Currents flowing parallel and in the same direction attract each other.*

2. *Currents flowing parallel and in opposite directions repel each other.*

3. *Currents making an angle with each other tend to become parallel and flow in the same direction.*

If we examine the direction of flow about a circular current forming an equivalent magnetic shell, we shall find that the current in every part of the circuit flows parallel and in the opposite direction to the current in the opposite diagonal part, Fig. 283. By Ampere's laws, therefore, the current in any part of an equivalent magnetic shell repels the current in the opposite part, so that if the conductor were made of elastic material, the shell would expand indefinitely. Possibly this may be seen more clearly by considering that the entire number of lines of force of the shell pass through the opening at the center; and since lines of force repel each other, the equivalent magnetic shell tends to expand, due to the crowding through its center of great numbers of these lines.

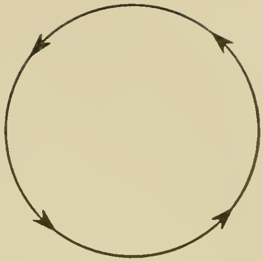

FIG. 283. The circle in which a current is flowing tends to expand.

481. The Solenoid. We have seen (p. 546) that a circular current has about it a field of force in every way similar to the field of force about a magnetic shell, and that such a current may in every way replace a magnetic shell, where the current coincides in position with the boundaries of the shell, and the lines of force due to the current are the same in number and direction as the lines of force which form the shell.

A series of circular currents, all flowing in the same direction with their planes parallel to one another, constitutes what is known as a *solenoid.* In practice, however, the solenoid is obtained not by having a series of strictly circular currents, but by winding the conductor through which the current is flowing into the form of a spiral, Fig. 284. This is equivalent to a series of circular currents and to a number of linear displacements, each equal to the distance between the turns of the spiral. Accordingly, if the straight conductor is carried back to the starting point, Fig. 285, this will counteract the field due to the linear displacements, and hence the field of force of the solenoid is strictly that due to the series of circular currents.

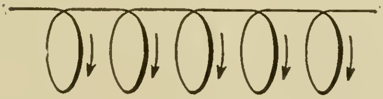

FIG. 284. Circular currents and linear displacements.

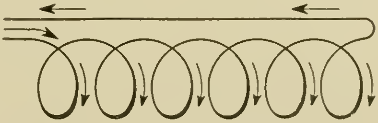

FIG. 285. A solenoid is strictly equivalent to a series of circular currents.

Through such a spiral conductor, we shall find, Fig. 286, that the currents an any side of the solenoid flow parallel to each other and in the same direction; and hence the field of force about any two adjacent turns of the conductor will be the same as that about any two currents flowing parallel and in the same direction. Accordingly, if we imagine the plane of the paper to be a longitudinal section through the solenoid, where the heavy line, Fig. 287, represents the solenoid itself, the currents in the successive turns will flow toward the observer along the upper side, and away from the observer along the lower side. Immediately about the wires the lines of force are circular in form; but at a greater distance, these lines, after passing around one conductor, include also the next and the next, and so on, and ultimately unite in one continuous line extending around the outside of the solenoid and returning through the center. Thus, we see that a solenoid is surrounded by a magnetic field, consisting of lines of force passing out at one end, extending around its entire length and entering the other end, only to return in closed curves to the starting point. But this, as we have seen (p. 423), is precisely the nature of the field of force about a magnet, where the lines of force emanate from the north pole of the magnet, pass in closed curves to the south pole, and at this point again enter the magnet, only to flow as lines of magnetization back through the body of the magnet from the south pole to the north.

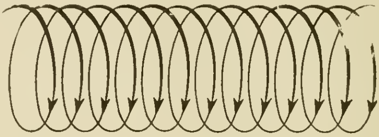

FIG. 286. In adjacent turns of the solenoid the currents flow parallel and in the same direction.

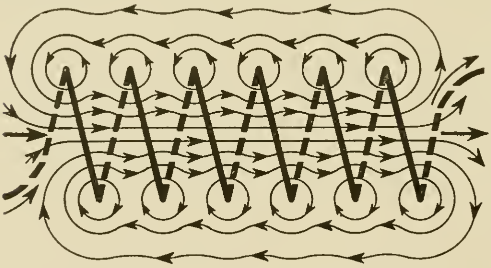

FIG. 287. The linear section of a solenoid, showing the field of force.

Since a solenoid has a field of force in every way similar to that about a magnet, a solenoid has all the properties of a magnet, and may in every instance replace the latter. A solenoid, accordingly, has two poles like those of a magnet, the one from which the lines of force emanate being the north pole, and the one into which they pass being the south pole. Inasmuch, then, as a solenoid has all the properties of a magnet, it will set itself north and south in the earth's magnetic field, the similar poles of two solenoids will repel each other while dissimilar poles will attract, the north pole of a solenoid will be repelled by the north pole of a magnet and attracted by the south pole, and so on. In fact, a solenoid behaves in every way like a magnet.

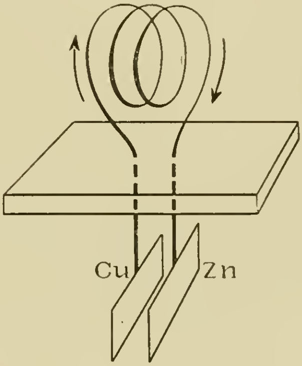

Fig. 288. De la Rive's floating battery.

A striking illustration of the fact that a solenoid in every case resembles a magnet, and that a very short solenoid behaves like a magnetic shell, is obtained by means of De la Rive's floating battery, Fig. 288. This consists of a plate of copper and a plate of zinc, connected above by a coil of wire wound in the form of a short, thick solenoid, the whole mounted on a cork so as to float in a solution of dilute sulphuric acid.

The plates of copper and zinc, thus being immersed in the dilute acid, form a battery from which the current flows through the coil of wire above. This then forms a little solenoid which, if left to itself, will set itself in the magnetic meridian so that its north pole will point north. If a magnet is approached, the little floating solenoid will be repelled or attracted according as the pole of the magnet brought near is of similar or of opposite nature to its own pole. Thus, if the north pole of the magnet is presented, the little floating solenoid will set itself so as to present its south pole to the magnet, and then attraction between the two will occur. The solenoid does not stop, however, when it reaches the north pole, but continues to move along the magnet, coming to rest only when it reaches the center, Fig. 289, after having slipped itself like a ring upon the magnet. When this has taken place, the little solenoid is in the position of maximum stability with respect to the magnet, that is, the relative potential energy between itself and the magnet has become a minimum. In this position, not only all the lines of force of the little solenoid pass through its interior space, but also all the lines of force of the magnet, as these flow from the south pole to the north through the body of the magnet.

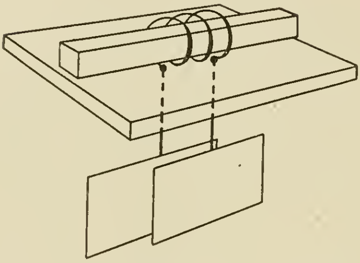

FIG. 289. The floating battery attracted by a magnet.

This is a beautiful illustration of the remarkable theorem announced some time ago (p. 433) that *a magnetic shell if placed in a magnetic field will tend so to move that the number of lines of force passing through it will be a maximum*, that is, that the potential energy between the magnetic field and the magnetic shell will be a minimum. Indeed, it is simply in accordance with this principle that the equivalent magnetic shell, or the solenoid, tends to set itself at right angles to the magnetic meridian of the earth's field, because in this position there pass through its area not only all the lines of force which it itself develops, but in addition just as many of the lines of force due to the earth's field as can be included within its area. Moreover, near the conductor through which the current is flowing, the lines of force in the earth's uniform field will, to some extent, be deflected, Fig. 290, so that as many of these lines as possible may pass through the area encompassed by the circular current.

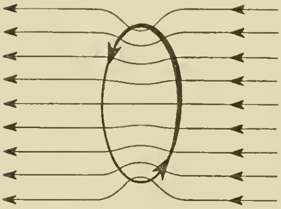

FIG. 290. A circular current in a uniform magnetic field.

Application has been made of this principle in the construction of the D'Arsonval galvanometer (p. 547). This instrument consists of a strong magnetic field, produced by permanent magnets, Fig. 291, and a delicately suspended coil of wire, through which is sent the current to be measured. In its normal condition, when no current flows, the plane of the coil is parallel to the lines of force of the field, and is maintained in this position by the feeble torison of the metallic suspending strips. When, however, a current is sent through the galvanometer, the suspended coil becomes a solenoid with a north and a south pole, and therefore tends to assume a direction at right angles to its former position, that is, with its poles adjacent to the opposite poles of the permanent magnet. In other words, it tends to set itself in the field of the permanent magnet so that not only its own lines of force but also those of the field in which it is placed may pass through it.

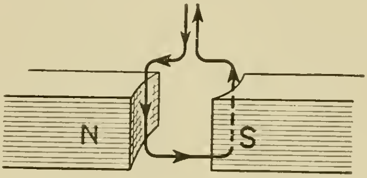

Fig. 291. The principle of the D'Arsonval galvanometer.

Another beautiful experiment illustrates the fact that a circular current will tend to set itself so as to include within its area the greatest possible number of lines of force. If a current is sent through a long, flexible conductor which is hanging beside a bar magnet and parallel to it, Fig. 292, the conductor will wrap itself a number of times in a spiral form around the magnet, thus forming approximately a series of circular currents, which constitute in reality a long solenoid. Through the center of this solenoid so produced the lines of force due to the current flow in the same direction as those of the magnet, so that the interior space of the solenoid includes the lines of force of the magnet as well as its own lines of force.

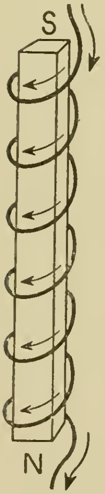

Fig. 292. A flexible conductor wraps itself about a magnet.

482. The Electromagnet. If we look down upon the *north pole of a solenoid,* the lines of force which pass through its interior will come directly towards us, and the north pole of a magnetic needle would likewise move in this way. By applying Ampere's rule for determining the direction of the current (p. 537), we shall find that the current which constitutes the solenoid, as we then view it, is flowing in the direction *opposite the motion of the hands of a watch.* In the same way we shall find that if we face the *south pole*, we perceive the current flowing around the solenoid *in the direction of the motion of the hands of a watch.* Therefore, a knowledge of the direction in which the current flows in the solenoid enables us at once to tell nature of the poles.

If a bar of soft iron is placed within the solenoid, it is then brought into the magnetic field which there exists, and so becomes magnetic. In other words, the lines of force, due to the current, which would otherwise flow through the interior space of the solenoid, now enter the iron and flow through its entire length. The presence of the lines of force within the iron causes each of the molecules to set itself parallel to these lines of force; or, in other words, the molecules of the iron are now arranged with their north poles pointing one way and their south poles pointing the other, and the bar of iron is therefore magnetized. It follows, accordingly, that a bar of iron placed within a solenoid around which a current is flowing becomes itself a magnet, and is known as an electromagnet, from the method of its formation. By this means artificial magnets may be made very much stronger than natural magnets, and, indeed, stronger than artificial magnets made in any other way.

483. Ampere's Theory of Magnetism. Since the poles of an electromagnet coincide in nature with those of the solenoid in which it is placed, we find that if we face the north pole of an electromagnet, Fig. 293, we see the current in the solenoid flowing in a direction *opposite* the motion of the hands of a watch, while if we face the south pole of an electromagnet, we find this current flowing *in* the direction of the motion of the hands of a watch. If the soft iron within the solenoid is replaced by steel, the steel likewise becomes magnetic, and retains its magnetism when it is removed from the solenoid.

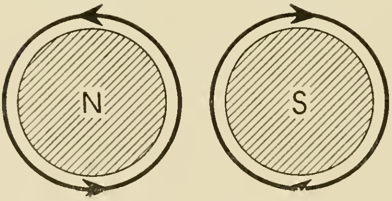

FIG. 293. The direction of the current about an electromagnet.

Inasmuch as the magnetic condition of the steel is originally produced by a current flowing through the solenoid around the magnet, *Ampere assumed that the magnetism of every permanent magnet is maintained by a current which continues to flow around the magnet.* In other words, Ampere believed every permanent magnet to be an electromagnet, and to retain its magnetic condition because of hypothetical currents flowing around it. Accordingly, if we face the *north* pole of a permanent magnet, we see these hypothetical currents, or the Amperian currents, as they are now called, flowing around the magnet in a direction *opposite* the motion of the hands of a watch, or counter-clockwise, Fig. 294; while if we look upon the *south* pole of a permanent magnet, we see these Amperian currents flowing around the magnet *in* the direction of the motion of the hands of a watch, or clockwise.

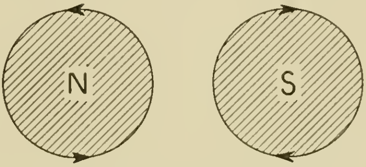

FIG. 294. The Amperian currents about a permanent magnet.

Moreover, according to Ampere's theory, a tiny current must flow around each molecule of a bar of iron or steel, since each molecule itself is believed to be a magnet. Around each molecule, therefore, there flows a minute Amperian current, counter-clockwise if we look down upon the north pole of the molecule, and clockwise if we face the south pole. If the bar is unmagnetized, Fig. 295, just as many of these little molecular currents flow in one direction as in the other, since there is no resultant magnetic effect. An examination, however, of these molecular currents reveals the fact that the condition of greatest molecular stability is produced when the currents in ad-

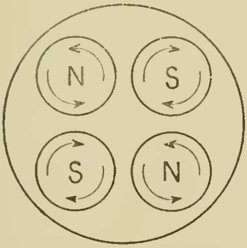

FIG. 295. Molecular Amperian currents in unmagnetized iron.

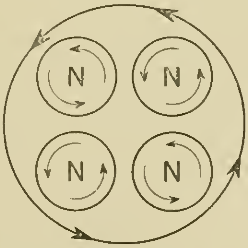

FIG. 296. Resultant Amperian current as seen from the north pole.

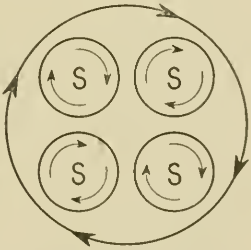

FIG. 297. Resultant Amperian current as seen from the south pole.

jacent molecules flow in opposite directions. It is then seen that if the molecular current flows clockwise around one molecule and counter-clockwise around a neighboring molecule, *these two currents flow parallel and in the same direction at those points where two molecules come together;* and since currents which are parallel and in the same direction attract each other, the two molecules are held in this position by the mutual attraction of these currents. When the bar is magnetized by being placed in a solenoid, the current in the solenoid compels all the molecular currents to flow in the same direction as itself, so that these molecules in which the molecular currents were formerly in the opposite direction are now turned over. According to this theory of Ampere, these tiny molecular currents continue to flow when the bar is withdrawn from the solenoid, and the condition of permanent magnetism which the bar now has is due to the persistence of these currents in flowing about the molecules.

The hypothesis of these molecular Amperian currents furnishes an explanation of the unstable condition of the molecules of a magnetized bar of iron or steel quite as well as does the theory which we have formerly considered. Although these small molecular currents flow in the same direction, that is, either clockwise or counter-clockwise, about all the molecules of a permanent magnet, it will be seen that in contiguous portions of adjacent molecules these Amperian currents flow parallel and in opposite directions. These currents, therefore, repel one another, and, accordingly, adjacent molecules likewise repel one another; but if the retentivity of the steel (p. 407) is sufficiently great, the molecules are nevertheless held in position, and the bar of steel becomes a permanent magnet. Around the outer edge of the bar, however, the molecular Amperian currents all flow in the same direction, and therefore all unite into one great resultant Amperian current which flows clockwise when viewed from the south pole of the magnet, Fig. 297, and counter-clockwise when seen from the north pole, Fig. 296.

484. The Modern Theory of Magnetism. For many years Ampere's theory of magnetism has been held to be most interesting, though highly improbable, inasmuch as no currents which could produce the molecular magnets were known to exist. Moreover, a magnetic theory based upon the attractions and repulsions of parallel currents would not have furnished an adequate explanation of magnetism, for the underlying cause of these attractions or repulsions could hardly be interpreted without the conception of the lines of force introduced later by Faraday. The discovery, however, of the electron, and a partial knowledge, at least, of the constitution of the atom, make it now seem probable that the theory of Ampere is, in the main, correct.

As has already been pointed out, (p. 340), an atom of matter is conceived at present to be a number of isolated electrons revolving with enormous velocities in regular orbits within the atom, either about a positive nucleus, or in a medium of "positive electrification". The electron, however, as it moves, thereby transports the elementary electrical charge of which it is composed; and the movement of this relatively enormously large quantity of electricity is what we now understand to be the electric current. *The revolution of the electrons in the iron atom, therefore, constitutes a current of electricity flowing around this atom;* and the series of circular currents thus produced convert this atom into an elementary electromagnet, with a north and a south pole, and a complete magnetic field.

Accordingly, after nearly a hundred years, we return for an explanation of magnetism to a modification of the Amperian theory. The current which Ampere assumed to flow about the atom is now found to exist in the revolution of the electrons within the atom; and, in addition, this current furnishes an explanation of the origin of the magnetic field of force, to which all magnetic phenomena are now believed to be due.

485. The Work Done by a Current in Moving Around a Unit Pole. Oersted's discovery that a magnetic needle moves when brought near an electric current shows that the needle is acted upon by a force which finds its origin in the current, or in the field of force produced by the current. We know, also that a free north pole will move in a circle about the current (p. 538), inasmuch as the lines of force about the current are complete circles. The force, however, which exists between the current and the needle must be a mutual one; and, therefore, just as a free north pole will move around the current if it is free to move and the current is fixed, so also will the current move in an opposite direction around a free north pole if the latter is fixed and the former is free to move. When we have succeeded in determining the work done by the current thus in making one complete revolution around a magnetic pole, it is obvious that the same expression will represent the work done by a free north pole in moving once around the current.

In determining this work, we may imagine the conductor through which the current is flowing to be bent into the form of a semicircle, and the unit pole to be placed at the center, Fig. 298. If, then, the current I flows through the semicircle from A to B, the north pole at the center will be urged away from the observer, while the conductor itself will tend to rotate out from the plane of the paper towards the observer about AB as an axis.

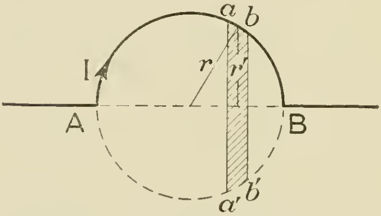

FIG. 298. A semicircular current revolving about a unit pole.

We have seen (p. 541) that the expression for the force between a current and a magnetic pole of strength m is

$$\frac{I\,L\,m}{r^2}.$$

In this case, however, since

$$m = 1,$$

the expression for the force becomes

$$\frac{I\,L}{r^2},$$

where L is the length of the conductor from A to B. If we now consider further that this semicircle is divided into elements of length, such as *ab*, the force which acts between each element of the current and the unit magnetic pole is

$$\frac{I}{r^2} \times a\,b,$$

and this force always remains at right angles to the current, that is, at right angles to the plane of the semicircle. If the semicircular current makes one complete revolution about the unit pole, the element *ab* will describe a zone on the surface of a sphere, represented by *abb′a′*. At the same time the force

$$\frac{I}{r^2} \times a\,b$$

which acts upon this element of length, will likewise be carried once around the circumference of a circle of radius r', and an amount of work equal to

$$\frac{I}{r^2} \times a\,b \times 2\,\pi\,r',$$

or

$$\frac{I}{r^2} \times \text{area of zone}$$

will be done. In the same way, each element of the semicircle will describe a zone, and will perform an amount of work equal to

$$\frac{I}{r^2} \times \text{area of the zone which that element describes.}$$

Accordingly, the total amount of work done by the entire semicircular current in making one complete revolution is

$$\text{work} = \frac{I}{r^2} \cdot \text{ sum of areas of all the zones,}$$

$$= \frac{I}{r^2} \text{ surface of sphere,}$$

$$= \frac{I}{r^2} \times 4\,\pi\,r^2,$$

$$= 4\,\pi\,I.$$

But the unit pole at the center of the semicircle sends out 4π lines of force (p. 429), and each of these lines of force is cut by the conductor in making one complete revolution. We may, therefore, consider that the work which is done by a current in revolving about a magnetic pole is really performed by the conductor cutting across the lines of force of the pole, and is equal in amount to the product of the strength of the current and the number of lines of force which are thus cut. Accordingly, in general, a current in a field of force tends to move in a direction at right angles to the lines of force, thereby cutting across the lines of force of the field; and the work done by the current thus in cutting across the lines of force of the field is equal to the product of the strength of the current flowing and the number of lines of force thus cut.

486. The Work Done in Carrying a Unit Pole Through the Loop Formed by a Circular Current. From the last paragraph it is evident that if the current I performs $4\pi I$ ergs of work in moving once around a unit pole, a unit pole will also perform $4\pi I$ ergs of work in moving once entirely around the current I. Accordingly, if this current flows in a circle, so as to form an equivalent magnetic shell, Fig. 299, a free north pole placed at P will move of itself entirely around the current, that is, it will move in such a way as

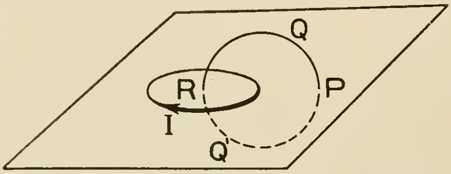

FIG. 299. A unit pole is carried through the loop formed by a circular current.

to pass through the loop formed by the current and back to its starting point along some path such as $PQRQ'$, and in so moving will perform

$$4\pi I \text{ ergs of work.}$$

If an outside force compels the unit pole to move once through the loop and back again to the starting point but in the opposite direction, that is, along the path $PQ'RQ$, the outside force will perform

$$4\pi I \text{ ergs of work.}$$

It is evident that the result will be the same if the unit pole is carried from any other point once through the loop and back again to the same point, and hence the proposition is generally true that the *work required to carry a unit magnetic pole once through the loop formed by a circular current and back again to the starting point is*

$$4\pi I \textit{ ergs.}$$

487. The Strength of the Field Inside a Solenoid. In a solenoid which consists of n turns of wire, through each of which the current I is flowing, we have the equivalent of n circular loops. Moreover, since it requires $4\pi I$ ergs of work to carry the unit pole once through a single loop and back again to the starting point, it will require

$$4\pi n I \text{ ergs}$$

to carry the unit magnetic pole entirely through the solenoid and back again to the starting point. Here I is measured in absolute units. If, however, as is more often the case, the current is measured in amperes (p. 482), the work required to carry the unit pole once entirely through the solenoid and back again to the starting point will be

$$\frac{4\pi n i}{10} \text{ ergs,}$$

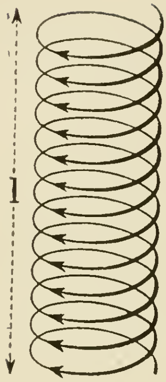

FIG. 300. A unit pole is carried through and around a solenoid.

where i expresses the current as measured in amperes.

We can, however, form another expression for the work done in carrying the unit pole once through and around the solenoid, and by equating these two quantities we can derive the expression for the intensity of the field within the solenoid. If we represent by H the strength of the field within the solenoid, Fig. 300, we mean that the unit pole placed in this field will be acted upon by a force of H dynes, so that we shall have to exert a constant force of H dynes in carrying the unit pole through the solenoid. If l represents the length of the solenoid, the work done in carrying the unit pole through the interior of the solenoid is

$$Hl \text{ ergs},$$

since (p. 117) work equals the product of the force acting and the distance through which it acts.

We must now remember that all the lines of force due to the current pass through the interior of the solenoid, while on the outside they pass through a region very many times larger, and hence the field of force on the outside of the solenoid is very weak indeed as compared with that upon the interior. Therefore, when a unit magnetic pole is carried entirely through a solenoid and around on the outside to the starting point, *practically all of the work herein involved is expended in carrying the pole through the interior of the solenoid,* while very little is required to carry it around to the starting point on the outside. If we neglect the latter small amount, we shall have, as the entire amount of work required to carry the unit pole once through the solenoid and around on the outside to the starting point, approximately,

$$Hl \text{ ergs}.$$

Accordingly, we have two expressions for the work required to carry a unit pole once through the solenoid and around to the starting point, hence,

$$H\,l = \frac{4\,\pi\,n\,i}{10},$$

and from this,

$$H = \frac{4\,\pi\,n\,i}{10l},$$

which is the expression for the strength of field inside the solenoid. This expression is true only upon the supposition that the solenoid is very long indeed as compared with its diameter, but for short solenoids it is wholly inadequate.

In addition to determining the strength of field, we are able also to determine *the total number of lines of force which pass through the solenoid.* When we say that the strength of field within the solenoid is H we mean that the unit pole placed there will be acted upon by H dynes of force. We have learned, however, (p. 428) that the expression for the strength of the field indicates also the number of lines of force which pass through one square centimeter of the field. Therefore, if the strength of the field inside the solenoid is $\frac{4\,\pi\,n\,i}{10l}$ we mean that $\frac{4\pi\,n\,i}{10l}$ lines of force pass through one square centimeter of the interior of the solenoid. If r represents the radius of the solenoid, the area of the solenoid is πr^2 square centimeters. Since

$$\frac{4\,\pi\,n\,i}{10l}$$

lines of force pass through one square centimeter and there are

$$\pi r^2$$

square centimeters in the area of the solenoid, there are, then,

$$\frac{4\,\pi\,n\,i}{10l} \times \pi\,r^3$$

or

$$\frac{4\pi^2 r^2 n i}{10l}$$

total lines of force passing through the solenoid.

From this expression for the total number of lines of force passing through the solenoid, we are further able to derive the expression for *the equivalent strength of pole of the solenoid.* We have learned (p. 429) that there are 4π lines of force emanating from a unit pole, and from a pole the strength of which is m there emanate $4\pi m$ lines of force. If, therefore, *M represents the equivalent strength of pole of the solenoid,* there emanate from this equivalent pole

$$4\pi M$$

lines of force. But we have just seen that the total number of lines of force passing out of the end of the solenoid is

$$\frac{4\pi^2 r^2 n i}{10l},$$

therefore

$$4\pi M = \frac{4\pi^2 r^2 n i}{10l},$$

from which

$$M = \frac{\pi r^2 n i}{10l}.$$

But πr^2 is the area of the cross-section of the solenoid; and therefore the expression for the strength of pole of the solenoid reduces to

$$M = \frac{\text{area}}{10} \times \frac{n i}{l}.$$

In all of the expressions here derived, it is to be observed that the factor ni occurs. Here n represents the number of turns of wire upon the solenoid, and i the current in amperes flowing through it. The product ni, therefore, is known as *the ampere-turns*, and this is the great factor upon which the magnetic effect, or the magnetizing power, of a solenoid depends. Since ampere-turns is essentially a product, precisely the same effect in every way will be produced by a current, for instance, of 2 amperes through 50 turns, or by 4 amperes through 25 turns, or by 5 amperes through 20 turns, or by 10 amperes through 10 turns, and so on. The result in each case is 100 ampere-turns, and the magnetic effect of each is the same. Moreover, since l represents the length of the solenoid, *the expression*

$$\frac{n\,i}{l}$$

represents the number of ampere-turns per unit of length. Introducing this expression into the three expressions just derived, we have

The strength of field within a solenoid

$$= \frac{4\,\pi}{10} \times \textit{ampere-turns per unit of length.}$$

The total number of lines of force passing through a solenoid

$$= \frac{4\,\pi^2 r^2}{10} \times \textit{ampere-turns per unit of length.}$$

The equivalent strength of pole of a solenoid

$$= \frac{\textit{Area}}{10} \times \textit{ampere-turns per unit of length.}$$

CHAPTER XLI.

THE MAGNETIZATION OF IRON.

488. The Permeability of Iron. If a bar of iron is placed within a solenoid, the lines of force there present enter the iron, and the latter becomes magnetic. The strength H of the field of force within the solenoid is called the *magnetizing force* of the solenoid, inasmuch as it is this field of force that causes the magnetic condition of the iron. As the lines of force enter the iron, the molecules arrange themselves parallel to these lines of force, and the iron thus becomes magnetized. Before the magnetization of the bar, however, the molecules of the iron are already tiny magnets, but are without systematic arrangement, so that in this condition they form groups among themselves in obedience to their own molecular attractions. The small field of force about each molecule, reacting upon similar adjacent fields, compels these molecules to assume positions of stability with respect to one another, so that the lines of force passing from molecule to molecule throughout the group may be as short as possible. As soon as the magnetizing force of the solenoid brings about an orderly arrangement of the molecules, these molecular fields unite into one great resultant field and the magnet is then surrounded by its field of force. The lines of force about a magnet, therefore, represent the sum of the lines of force about as many of the molecular magnets as lie parallel to one another and in the same direction.

Accordingly, in an electromagnet there are present really two sets of lines of force, or at least lines of force which have their origins in two different sources:

1. Those which constitute the field of the solenoid, and which cause the magnetization of the iron within.
2. Those which belong to the iron core in virtue of the fact that this is now a magnet and hence is surrounded by its own field of force.

Both of these sets of lines of force pass through the body of the iron which fills the interior of the solenoid, and hence the number of lines of force passing through a solenoid is enormously increased by placing a bar of iron in its center.

It is now possible to regard this phenomenon in two entirely different ways. From the first point of view we may think of the two sets of lines of force as maintaining their individuality and remaining distinct throughout their entire circuit. Just as H denotes the number of lines of force that pass through one square centimeter of the solenoid before the iron is present, so B is used to represent the total number of lines of force per square centimeter that pass through the iron when the interior of the solenoid is filled with iron instead of air. Therefore, B is always a composite quantity, being the resultant of H and the number of lines of force per square centimeter developed within the iron when the latter becomes a magnet.

When the iron within the solenoid becomes magnetized, it develops poles of strength m, and from the north pole of this magnet there emanate $4\pi m$ lines of force (p. 429). These lines of force pass in the curves of lines of force to the south pole, and flow back as lines of magnetization (p. 423) to the north pole through the body of the magnet. If the magnet is a square centimeters in cross section, $4\pi \frac{m}{a}$ lines of force emanate from each square centimeter of the polar surface, and $4\pi \frac{m}{a}$ lines of magnetization also flow back from the south pole to the north through the body of the magnet. Therefore, each square centimeter in the cross section of the iron within the solenoid carries H lines of force due to the field of the solenoid, and also $4\pi \frac{m}{a}$ lines of force due to the magnetization of the iron, so that

$$B = H + 4\pi \frac{m}{a}.$$

But (p. 416) we have learned that the expression $\frac{m}{a}$ is the quantity known as the *intensity of the magnetization*, I, and hence

$$B = H + 4\pi I.$$

From the other point of view we may regard the iron as an extremely good conductor of lines of force, or, to use Kelvin's expression, as extremely *permeable* to lines of force. Lines of force crowd into the iron because this medium offers a very much easier path for lines of force than does the air, so that the number of lines of force is enormously increased by allowing them to pass through the iron. Accordingly, a given magnetizing force due to the solenoid will develop a very much larger number of lines of force in the iron placed within, and the ratio of the number of lines of force per square centimeter developed within the iron to the number per square centimeter originating in the solenoid is a measure of *the conductivity of the iron for lines of force, or the permeability.* If μ denotes this factor of increase in the number of lines of force in the iron, that is, if μ denotes the ratio by which B is larger than H, or the permeability, then

$$B = \mu H,$$

and

$$\mu = \frac{B}{H}.$$

Permeability, therefore, may be regarded as in every way analogous to *conductivity for lines of force*, and may be defined as *the ratio of the number of lines of force per square centimeter developed within the iron to the strength of the field in which this iron is placed.* In other words, μ is defined as the ratio of B to H.

Of the two views of magnetization here presented, the latter is possibly the more direct and is the one more commonly employed. This is true, because in practice the object is mainly to ascertain the number of lines of force developed within the iron, without raising the question as to the source of particular lines, or their division into groups. The former view, however, is far more instructive, for here we trace the origin of the lines of force which appear in the iron, and are offered a full and complete explanation of the forces of magnetization, certainly so far as the phenomenon of magnetization can be explained.

The quantity here designated as B is known as *the induction* in the iron, and the lines of force, so long as they remain in the iron, are spoken of as lines of induction. It is evident that the degree of magnetization depends upon the number of these lines of induction, the greater the induction the higher the degree of magnetization. *Permeability, therefore, may be defined as the ratio of induction produced in the iron to the magnetizing force of the field.*

The magnetization of iron, however, is complicated by the fact that *the ratio of B to H, or the permeability, is not constant, but depends upon the strength of the field H, upon the quality of the iron, and upon the previous history of this iron as regards magnetization.*

489. The Curve of Magnetization of Iron. An excellent idea of the value of π for different strengths of field may be obtained from the curve known as *the curve of magnetization.* In this curve, Fig. 301, the different values of H are plotted as abscissae, while the corresponding values of B are plotted as ordinates. We thus obtain a curve from which the value of μ may be determined for each corresponding value of H.

As we plot such a curve, we find that for very small values of the magnetizing force very little induction in the iron is produced, therefore the first part of the curve is flat and does not depart very much from the horizontal axis. Soon, however, as the magnetizing force is increased, the induction in the iron increases almost in direct proportion to the magnetizing force, so that there is a portion of the curve that slopes rapidly upward and is nearly a straight line. The height to which the straight portion rises

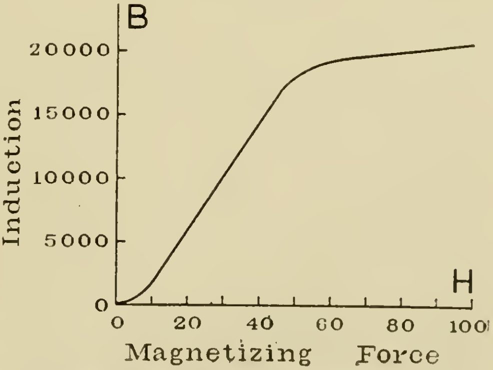

FIG. 301. The curve of magnetization of iron.

varies for different kinds of iron. For good specimens of soft wrought iron, *B*, may here reach a value as high as 18,000, while the corresponding value of *H* is about 50. The curve then makes a gradual bend to the right, and continues to rise only very slightly, being nearly parallel to the horizontal axis, so that a further increase in the magnetizing force produces only very slight increments in the induction in the iron.

An investigation of the molecular changes which take place in the iron during magnetization furnishes an explanation of the facts set forth in this curve of magnetization. When a very weak current is sent through the solenoid, that is, when the magnetizing force is very small, we have seen that a very feeble induction is produced in the iron, so that the first portion of the curve is very low and flat, and is nearly parallel to the horizontal axis. This is interpreted as meaning that the molecules of the unmagnetized iron have settled into positions of maximum stability, and form groups or chains among themselves, the north pole of one molecule attracting and holding in position the south pole of the one adjacent. When a feeble magnetizing force is applied, certain molecules are slightly displaced from their former, more stable positions, although the groups are not yet broken up. As soon as the magnetizing force is removed, these molecules spring back again into their former relations to one another, and so contribute nothing to the permanent, or residual, magnetism of the bar. This constitutes the first stage in the process of magnetization.

After these first small deviations in the molecular groupings have been effected, some of the molecules forming a cluster, or chain, are then just on the point of turning over. An increase, therefore, in the magnetizing force compels at first a few, and then more and more, of the molecules to set themselves in the direction of the lines of force, so that during this, which is the second stage in the process of magnetization, the induction increases almost exactly in proportion to the magnetizing force, and therefore this part of the curve is very nearly a straight line. During this stage of the process, a larger and larger number of the molecules are setting themselves parallel to one another, so that the induction which is produced may be taken as indicative of the relative number of the molecules that are thus arranged in parallel positions.

After a magnetizing force of about 50 has been reached, it is found that nearly all the molecules that can be readily turned about by a moderate magnetizing force have fallen into position parallel to those which have thus been turned previously. For a little distance on the curve, therefore, it is found that an increase in the magnetizing force turns into position only a small number of molecules, because nearly all have already set themselves parallel to one another, and those which are not thus turned are very rigidly bound to their neighbors. It is on this account that the bend of the curve is produced. *At this point, therefore, it is certain that practically all the molecules are parallel to one another and pointing in the same direction;* and as it is impossible to make these molecules any more than parallel, *it is likewise impossible to increase the induction beyond this point.* Therefore, at the bend of the curve the magnet is said to be *saturated,* because at this point the molecules are parallel to one another.

The third step in the process of magnetization lies beyond this point. Here it is found that enormous increases in the magnetizing force are successful in turning into position only occasional molecules, or possibly by forcing those which have turned into position to set themselves more strictly parallel to the lines of force. The increase in the magnetism produced, however, is only very slight indeed; and in practice it is found wholly useless to carry the magnetization beyond the point of saturation.

Ewing has devised a magnetic model consisting of a number of little magnetic needles mounted within a solenoid. The little needles represent the molecules of an unmagnetized bar of iron, and point in all possible directions, forming long chains or groups, Fig. 302, while the solenoid represents the magnetizing coil by which this is to be magnetized. A very feeble current sent through the solenoid will cause certain lines or groups of molecules to turn only slightly, Fig. 303, without, however, breaking up or interrupting the particular molecular chains of which they are members, thus representing the first stage in the process of magnetization. After this, as a stronger current is sent through the solenoid, it is found that the needles turn into position nearly in proportion to the increase in the current, until finally they are practically all turned and lie parallel to one another, Fig. 304, though, owing to their mutual attractions, they are not absolutely parallel to the lines of force of the solenoid. This illustrates the second stage, and the "magnet" is now saturated. As the current is still further increased, the main effect to be observed is a slight movement of all the magnetic needles, Fig. 305, as they set themselves more nearly exactly parallel to the lines of force of the solenoid. This illustrates the third stage of the process, where a very great increase in the magnetizing force produces only a very slight increase indeed in the strength of the magnet.

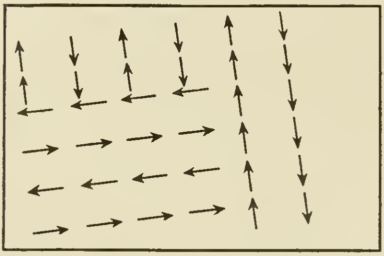
FIG. 302. Chains and groups of molecules in unmagnetized iron.

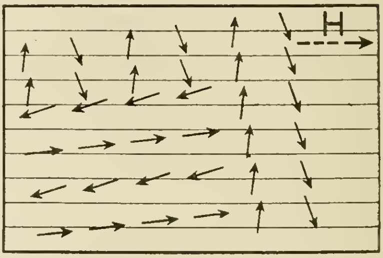

FIG. 303. First stage of magnetization—the molecular chains are not yet broken.

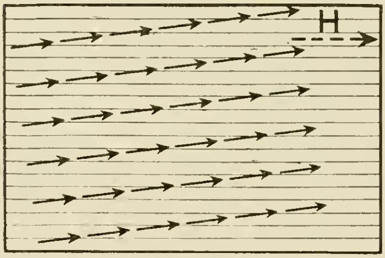

FIG. 304. Second stage of magnetization—the molecules are pointing in one direction.

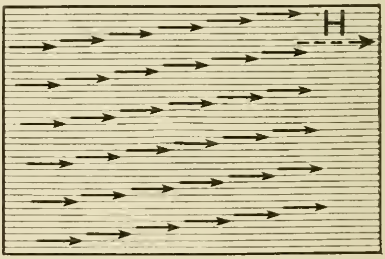

FIG. 305. Third stage of magnetization—the molecules are parallel to the lines of force of the field.

496. Hysteresis. If a bar or iron is subjected to an increasing magnetizing force H, the induction B will increase in the iron until the latter becomes saturated, and beyond this point B will increase but very little, even though H may be very great.

The curve which shows the magnitude of B for given values of H is the curve of magnetization which has just been described. If an unmagnetized bar of iron is subjected to this process of magnetization and the results plotted on axes of H and B, we shall find as before that as the magnetizing force H increases, the value of B increases, and the curve of magnetization Oa, Fig. 306, is obtained.

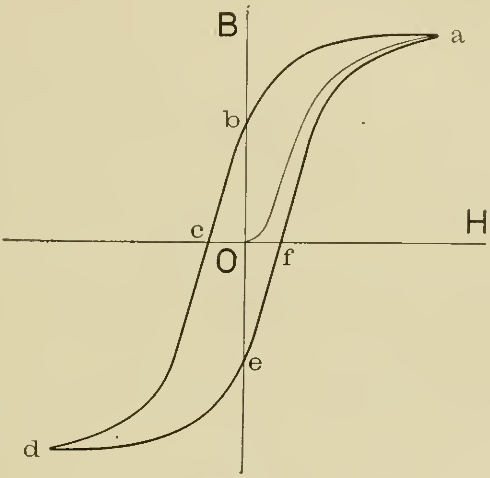

FIG. 306. The hysteresis loop.

After the iron has thus been magnetized, it will be found that, if the magnetizing force is decreased, the value of B will also decrease. If the molecules of the iron were perfectly free to move, B, would always have the same value for a given magnetizing force, whether the latter were increasing or diminishing. But even in the softest iron, the molecules, due to their mutual polar attractions for one another, Fig. 304, have a certain rigidity of position which prevents them from turning around with perfect freedom. After the iron has been fully saturated, the molecules, due to these attractions, tend to some extent to stay in this position of parallelism, so that for the same value of H the value of B is greater on the descending curve than on the ascending one. For decreasing values of H, therefore, the curve of magnetization does not coincide with the curve obtained for increasing values of H, but lies considerably higher than the latter, thus forming the portion ab of the curve. Indeed, when the magnetizing force has been reduced to zero, a large number of the molecules of the iron still stay in position, producing an induction indicated by the distance Ob. This distance Ob, then, represents the magnetism in the iron after the latter has been magnetized and the magnetizing force entirely removed, that is, *Ob* represents the residual magnetism (p. 414). It is evident that *b* will lie nearer the point *O* for soft iron than for very hard iron or steel.

If the current through the magnetizing solenoid is now reversed, the magnetizing force will also be reversed, and will tend to cause the molecules to set themselves in the opposite direction. But the iron has now an amount of residual magnetism represented by *Ob;* hence a magnetizing force equal to *Oc* must be developed in the opposite direction, in order to destroy this residual magnetism and reduce the iron once more to its completely unmagnetized condition. *Oc* then indicates the degree of stability of the permanent magnetism, and is called the *coercive force* of the iron, for this is the magnetizing force that must be applied in the opposite direction to overcome the molecular polar attractions, that is, to destroy the residual magnetism. As the magnetizing force *H* is still further increased, the magnetism of the iron now increases, of course in the opposite sense to its former magnetization; until the point of saturation is reached at *d*. If the magnetizing force is now once more decreased, the magnetism of the iron will again lag behind, so that, when *H* has again been reduced to zero, there still remains an amount of residual magnetism in the iron represented by *Oe*. The current in the magnetizing solenoid must, therefore, again be reversed, and the magnetizing force brought to the value *Of* before the iron loses its residual magnetism. Here again the magnetizing force *Of* represents the coercive force of the iron. It is to be observed, however, that although the iron is now once more in its original unmagnetized state, the curve does not close, since the magnetizing force *Of* still acts. *H*, therefore, must again be increased, and this will cause *B* to increase in value until the point of saturation *a* is once more reached, when the curve closes, and the cycle is complete.

The iron has now been magnetized to saturation, demagnetized, magnetized to saturation in the opposite direction, demagnetized, and again magnetized to saturation in its original direction, so that the values of *B* and *H* coincide with their former values. An entire cycle of operations has therefore been completed, and a closed, loop-shaped curve *a b c d e f* has been traced. The lagging of the induction behind the magnetizing force, due to the molecular polar attractions, is termed *hysteresis* (to lag behind), and this loop, through one cycle of magnetization, is called the hysteresis loop. The softer the iron, that is, the freer molecules, the more nearly will the two sides of the loop come together, while the harder the iron, or the more rigidly held the molecules, the greater is the residual magnetism, the greater also is the coercive force, and the more widely are the two sides of the loop separated.

491. The Loss of Energy Due to Hysteresis. The rigidity of the molecules, which has been mentioned above, must be construed as due not at all to forces of the nature of friction, but solely to the polar attractions which the molecules have for one another. On account of these polar attractions, the molecules of the unmagnetized iron form groups of maximum stability, so that work must be done by the magnetizing force in magnetizing the iron, that is, in turning these molecules around and making them lie parallel to one another. After this has been accomplished and the magnet is thus saturated, if the magnetizing force is then decreased to zero, a few of the molecules return to their undisturbed positions, and in so doing give back the energy which was expended upon them in breaking up the original groups. The larger number, however, remain in their condition of parallelism, thus causing the descending part of the curve to deviate from the ascending portion; and as these molecules thus lag behind the magnetizing force, they occasion the phenomenon of hysteresis, and give rise to the residual magnetism which the iron then possesses.

Although thus held parallel to one another, these molecules are in a relatively unstable position, so that a comparatively small amount of energy must be expended upon them to bring them again into their former groups, or to reduce the iron once more to its unmagnetized condition. Therefore, when an entire cycle of magnetization has been completed, and the iron is again brought back to a given magnetic condition, it is obvious that more energy has been expended in breaking up the original stable molecular groups when the iron was magnetized, than has been returned by the molecules in once more forming these groups during the process of demagnetization. In other words, a perfectly definite resultant expenditure of energy is required to carry a given sample of iron through a complete cycle of magnetization. Moreover, since the magnetic condition of the iron is the same before and after the completion of the cycle, this energy cannot be stored up in the iron, and the question may fairly be raised as to the form in which it then exists.

The answer to this question may be obtained by considering that, when the magnetizing force is applied, the molecules turn, at first slightly, and then more and more, without, however, destroying their original groups, Fig. 303, until this force becomes sufficiently intense to cause these particles to break away from their stable positions and assume the parallel condition. The potential energy which has thus been stored up in the molecules before they yield becomes kinetic when the rupture is finally effected, so that they fly past their new positions of equilibrium, and oscillate back and forth many times about these points before coming to rest. The vibration of each molecule causes similar vibrations to take place among the adjacent molecules; and as the kinetic energy of the molecules is thus increased, the temperature of the iron is thereby raised.

Accordingly, the resultant energy which is expended in carrying a sample of iron through a complete cycle of magnetization is wasted as regards effective results, and is dissipated in the form of heat. The amount of energy thus lost depends, therefore, upon the rigidity of the molecules, that is, upon the hardness of the iron; and this rigidity, as we have seen, is the cause of the lagging of the induction behind the magnetizing force, which gives rise to hysteresis. The softer the iron, the more nearly does the induction follow the magnetizing force, and the narrower is the loop; while the harder the iron, the greater is the coercive force required to destroy the residual magnetism, and the wider is the loop. Thus, the area of the hysteresis curve is indicative of the hardness of the iron, and of the energy losses which occur during the process of magnetization and demagnetization.

The form and area of the hysteresis loop are consequently of the greatest service in ascertaining the quality and the characteristics of a sample of iron; for this loop indicates at once the permeability, the retentivity, and coercive force of the iron, while its area is a measure of the energy which has been wasted, or dissipated in heat, in carrying this particular kind of iron through a magnetization cycle. Therefore, according to the purpose for which it is to be employed, the iron should be selected with reference to its qualities as shown by the hysteresis loop. For the armatures of dynamos (p. 603), and for the cores of transformers, (p. 599), where the iron must be rapidly magnetized and demagnetized, a good quality of soft wrought iron should be used, since this has a high permeability, a low coercive force, and a hysteresis loop so slender that it indicates a small energy-loss by heating. For permanent magnets, however, glass-hard tungsten steel or, still better, molybdenum steel, should be selected; for, while these have a lower permeability, they also have hysteresis loops which are very broad and thick, thereby showing that each of these hardened steels has an extremely great coercive force, which is the quality essential for maintaining the strength of a permanent magnet.

CHAPTER XLII.

ELECTROMAGNETIC INDUCTION.

492. Electromotive Force Generated by Cutting Lines of Force. In a former article (p. 566) it was shown that $4\pi I$ ergs of work are done when a semicircular current makes one complete revolution about a unit pole at its center, and that this work is done by the current in cutting the 4π lines of force which this unit pole sends out. If R represents the resistance between A and B, Fig. 307, and E is the difference of potential which exists between these points when the conductor is at rest, and if, furthermore, the semicircular current makes one complete

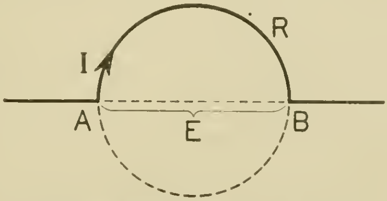

FIG. 307. A semicircular current revolving about a unit pole.

revolution in t seconds, then by the law of the conservation of energy, the energy of the current must equal that consumed as heat in the conductor plus the amount expended by the current in doing external work. Accordingly (p. 520),

$$I\,E\,t = I^2\,R\,t + 4\pi\,I,$$

from which

$$I = \frac{Et - 4\pi}{Rt},$$

or

$$I = \frac{E - \frac{4\pi}{t}}{R},$$

The expression is in the form of the equation representing the flow of current by Ohm's law, where the numerator of the fraction is the total electromotive force acting, and the denominator is the resistance through which the current flows. It is thus seen that *the expression*

$$\frac{4\pi}{t}$$

is of the nature of an electromotive force, and that this electromotive force has been generated by the conductor cutting across the lines of force of the magnetic pole. Furthermore,

$$\frac{4\pi}{t}$$

likewise expresses the *rate* at which the 4π lines of force which emanate from the unit pole are cut by the conductor. We have, therefore, the all-important law that *an electromotive force is generated by a conductor cutting across the lines of force of a magnetic field, and that this electromotive force is measured by the rate at which these lines of force are cut by the conductor.*

Since this electromotive force does not depend upon the difference of potential already existing, or upon the current flowing in the conductor, the principle is one of universal application, that *any conductor cutting across the lines of force of any magnetic field develops an electromotive force within the conductor, and the measure of this electromotive force is the rate at which these lines of force of the field are cut by the conductor.* Accordingly, if N lines of force are cut by a conductor in t seconds, an electromotive force e is generated, where

$$e : \frac{N}{t},$$

or

$$e = k\,\frac{N}{t},$$

where k is the factor of proportionality.

If we agree that e shall be taken as unity when N and t are each equal to unity then

$$k = 1,$$

and the equation becomes

$$e = \frac{N}{t},$$

The fact, however, that k is equal to unity in this expression limits the definition of the unit electromotive force to the conditions prescribed above. Accordingly, we learn that *the unit electromotive force is generated when lines of force are cut by a conductor at the rate of 1 line per second.* Moreover, since t is expressed in C. G. S. units, and N is measured in electromagnetic units, e is thereby determined in electromagnetic units. The unit electromotive force defined above, is therefore, the same electromagnetic unit of electromotive force which has been previously defined (p. 497) in terms of other C. G. S. units.

Inasmuch as an electromotive force of 1 electromagnetic unit is thus generated by cutting lines of force at the rate of 1 line per second, it follows that *an electromotive force of 1 volt, or 10^8 electromagnetic units will be generated by a conductor which cuts lines of force at the rate of 10^8 or 100,000,000 lines per second.*

If the two ends of the conductor which cuts across the lines of force are not joined, that is, if the conductor forms simply an open circuit, the electromotive force thus generated will cause a difference of potential to be established between the two ends of the conductor. If, however, these ends are brought together, thereby forming a closed circuit, this electromotive force will cause a current to flow through the circuit so formed. Indeed, this is the modern method by which an electric current is generated.

493. A Working Hypothesis of Electromagnetic Induction. From the preceding paragraph, it is evident that the two essential conditions for the production of a current by this means are, first, the existence of a magnetic field of force, and, second, the motion of a conductor forming part of a closed circuit across the lines of force of this field. Moreover, since the flow of current represents a certain amount of energy (p. 496), it is also evident that, in order to supply this energy, work must be done in carrying the conductor across these lines of force.

The current which is hereby generated is surrounded, as is every other current, by its circular magnetic field; and as this field is invariably the accompaniment of the current, we accept without further question the fact that a current is flowing, whenever this field is produced or detected. This circular field is intimately connected with the lines of force of the field through which the conductor is moved, and its direction is entirely dependent upon the direction of these lines of force, as well as the direction of the motion of the conductor.

As a purely working hypothesis of the production of this circular field, and as a means of forming a picture which will enable us to connect the direction of flow of the current with the motion of the conductor and with the lines of force of the field, we may imagine that *the circular field about the current is a portion of the original magnetic field which has been* *wrapped about the conductor,* as the latter moved through this field. This conception will be the easier to form from the fact that the lines of force about the current are of the *same nature* as those of the original field, and, furthermore, that *their direction is the same as if the lines of force of the field had been elastic cords, and had been wrapped about the conductor in its motion across them.* In a similar way, by analogy, we may explain the *energy* of the current as the *work* which is required to distend the elastic, rubber-like lines of force of the field and wrap them as circular lines of force about the conductor.

Since the flow of current through a conductor invariably accompanies the circular field of force about it, and since this circular field is produced when a conductor sweeps across the lines of force of a magnetic field, we may imagine that the current which flows is a direct consequence of the formation about the conductor of this circular field of force. So consistently does the hypothesis herewith presented interpret the production of this field, that we shall use it henceforth *as if* it were a demonstrated fact, claiming for it, however, nothing more than a convenient means of determining in advance the direction of the current generated whenever a conductor is carried across the lines of force of a magnetic field.

494. **Currents Produced by Induction.** Accordingly, in the generation of a current by this method, we must keep constantly before us the fact that an electric current in a conductor is invariably accompanied by a field of force around the conductor, and that the presence of this circular field of force is conclusive evidence that a current of electricity is flowing. It will also be necessary continually to bear in mind the properties which we have attributed to lines of force (p. 424), namely, they are always closed curves, they behave like tightly stretched rubber cords, they tend always to contract, they never intersect, they repel one another laterally.

With these facts before us, we will now consider the effect produced when a portion of a closed conductor, in which no current is flowing, moves across the lines of force of a magnetic field. Let us assume that we are dealing with the earth's uniform field, and that we have a closed coil of wire, one part of which we can carry horizontally across these lines of force. Let the black dots, Fig. 308, represents the cross-section of the conductor in its various positions as it is carried to the right horizontally across the lines of force of the field.

When the conductor approaches a line of force, this elastic line is not cut at once, but bends, so that a slight depression is at first produced, as at *A*. As the motion of the conductor continues, this indentation of the line increases, as at *B*, and the line wraps itself partly around the conductor. A further motion of the conductor in the same direction distends the line of force to a still greater extent, as at *C*, and a circular line of force about the conductor is very nearly complete. When, finally, the conductor has been carried entirely across this line of force, a total rupture of the line is effected, the two broken ends unite, and the circle of the line of force about the wire is now complete, as at *D*.

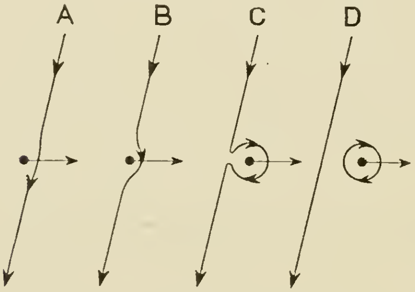

Fig. 308. Current flowing from the observer, induced by cutting lines of force.

Therefore, by the mechanical process of carrying a portion of a closed conductor across the lines of force of a magnetic field, an electromotive force is generated within the conductor, a circular field of force is wrapped about the conductor, and, accordingly, a current of electricity is developed flowing in the conductor. Moreover, the direction in which this current flows bears a perfectly definite relation to the direction of the motion of the conductor across the lines of force of the field. Inasmuch as the observer sees the lines of force in the figure flowing about the wire at *D* in the direction in which a screw would be turned (clockwise), it is obvious that the current in the conductor flows away from the observer, or into the plane of the paper.

If the conductor is carried to the left across the same field of force, Fig. 309, the lines of force of the field will be wrapped about it in the opposite direction. The same series of operations, therefore, will be carried out in the reverse order, as at *A′*, *B′*, *C′*, *D′*, until finally the circular line of force about the wire is complete, as at D'. Here, however, the lines of force pass around the wire counter-clockwise, hence the current which is developed is flowing towards the observer, or out from the plane of the paper. Therefore, an electric current may be generated in a closed conductor simply by carrying a part of the conductor across the lines of force of a magnetic field, the direction of the current thus developed depending upon the direction of the motion of the conductor with respect to the lines of force of the field.

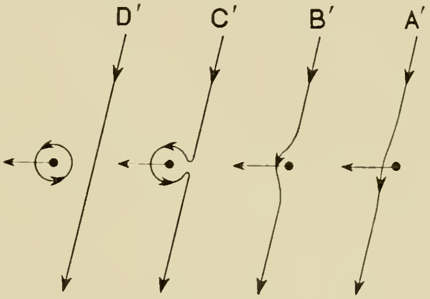

FIG. 309. Current flowing towards the observer, induced by cutting lines of force.

While the above analysis refers to the various steps in the process of cutting a single line of force, Figs. 310 and 311 give more nearly the conditions as they exist when a conductor is swept across the lines of force of a uniform magnetic field, and show very clearly the characteristics of elasticity and flexibility peculiar to magnetic lines of force.

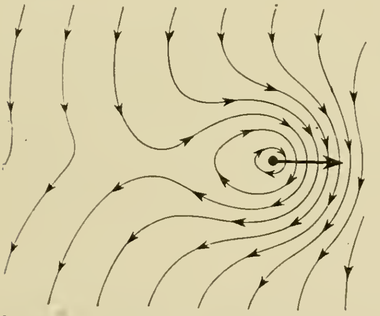

FIG. 310. Conductor sweeping to the right across the earth's field of force.

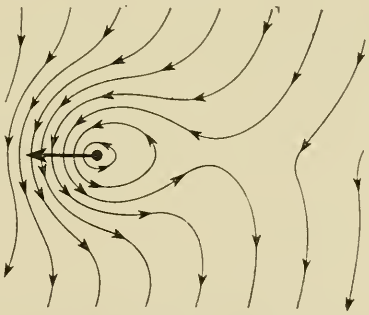

FIG. 311. Conductor sweeping to the left across the earth's field of force.

In this connection it is to be noticed that, if a part of a conductor cuts across the lines of force of the field, *a change occurs in the total number of lines of force that pass through the area encompassed by the closed circuit*. Accordingly, the above phenomena may be expressed by saying that an *increase* in the number of lines of force passing through a circuit will generate a current in the conductor flowing in *one direction*, and a *decrease* in the number of lines of force passing through the circuit will develop a current flowing in the *opposite direction.* Therefore, *any change in the number of lines of force passing through the area encompassed by a closed circuit will generate a current flowing through the conductor.* Since the current which is generated in a closed circuit depends upon the *change* of the number of lines of force passing through the circuit, we see that a closed coil of wire, moved parallel to itself in a uniform magnetic field, as at *A*, *B*, *C*, Fig. 312, always includes the same number of lines of force, and hence no current is developed. If, however, the plane of the coil changes its inclination, as, for instance, if it turns into the positions *D*, *E*, *F*, the number of lines of force passing through the circuit is changed, an electromotive force is created, and a current flows.

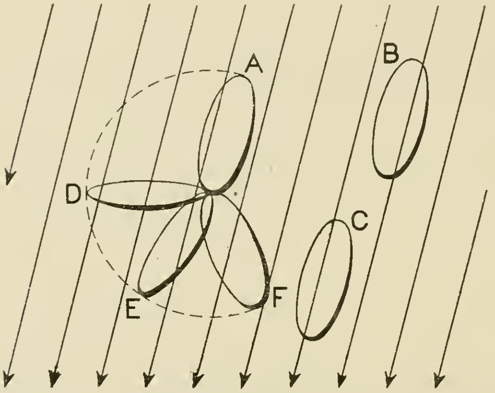

FIG. 312. Current induced by increasing or decreasing the number of lines of force passing through a closed circuit.

In all instances where a current is thus developed by cutting lines of force in a magnetic field, the current so generated is called an *induced* current, and the process of producing such a current is known as *induction.* It is by induction that nearly all the currents in use at the present time are generated; and, indeed, the principle of the dynamo is simply that of a conductor cutting across the lines of force of a magnetic field, or, rather, that of a closed circuit increasing and decreasing the number of lines of force which pass through it. In this connection, one very important fact must be borne in mind, namely, that *an induced current flows only while the lines of force of the field are being cut, and therefore, the current lasts only as long as the motion continues.*

495. Lenz's Law. An examination of the conditions under which an induced current is generated reveals the fact that the *mechanical energy,* expended in carrying a conductor across the lines of force of the field, *is consumed in overcoming the mutual repulsion between these lines of force and the lines of force wrapped about the conductor.* Thus, for instance, at *D,* Fig. 308, the conductor has been carried to the right across the lines of force of the field, and the current generated flows away from the observer. It is to be noted, however, that on the side of the conductor which is an advance, the circular lines of force, in passing around the conductor, lie in the same direction as the lines of force of the field and parallel to them. Since lines of force which are parallel and in the same direction repel one another, it is obvious that, *on the side of the conductor towards which the motion is taking place, the lines of force of the field repel those due to the current, so that the conductor tends to move in the opposite direction.* In the above illustration, we see that, while the current is produced by moving the conductor to the right, the reaction between the lines of force of the field and those about the conductor is such as to make the conductor *tend* to move to the left. It follows, therefore. that *the conductor in which the current is induced tends to move in a direction opposite to the motion by which the current is generated.*

This is but an illustration of a law which has the widest application in the production of induced currents, and a law to which there are no exceptions, namely, that *the current which is produced by induction always flows in such a direction as to oppose the motion which produces it.* This is known as *Lenz's law* and is of the most frequent use in determining the direction of the induced current.

496. The Production of Induced Currents. Illustrations of the generation of induced currents abound whenever a conductor and lines of force are moved with reference to one another. The same effect will be produced whether the conductor is moved across the lines of force of the field, or the lines of the field are carried across the conductor. In either case, these lines of force are wrapped around the conductor, and this is equivalent to generating a current in the conductor.

One of the earliest devices for producing a continuous current by induction is Faraday's disc dynamo. This consists of a copper disc, Fig. 313, rotating between the poles of a horse-shoe magnet, thus cutting the lines of force of the magnet as they extend from the north pole to the south pole through the copper. The part of the disc directly between the poles of the magnet is the conductor, and this is carried across the lines of force of the field. A current is thus generated in the copper disc flowing from the axis to the circumference, or from the circumference to the axis, according to the direction of rotation of the disc. If, therefore, contact is made with the edge of the disc, and a wire connects this contact with the axis, a current will flow through the circuit so formed, and will continue to flow as long as the rotation of the disc is maintained.

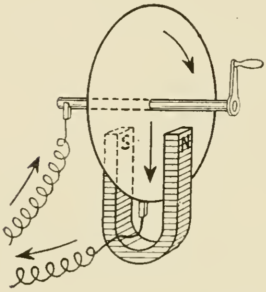

Fig. 313. Faraday's disc dynamo.

Another illustration of the production of an induced current may be shown when a conductor is suddenly moved towards a magnet, thus cutting the lines of force of the surrounding magnetic field. In Fig. 314, the black dot represents the cross-section of a conductor which is being carried in the direction of the arrow towards the magnet. The elastic lines of force are bent around the conductor, as the latter approaches, until finally entire circles of these lines are thus wrapped about it. Under the conditions shown in the figure, it is evident that the current generated in the conductor flows towards the observer. Attention is called to the fact that on the side towards the magnet the lines of force around the conductor are in the same direction as the lines of force of the field, and

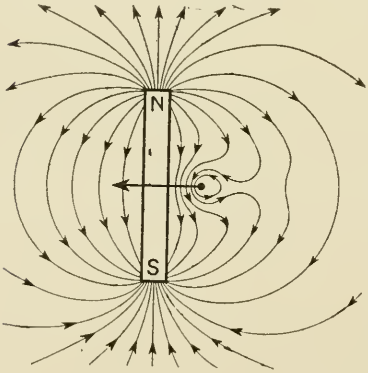

Fig. 314. Current induced by moving a conductor towards a magnet.

hence these two tend to repel. *But the current is generated in the conductor only by bringing the two together.* Hence, we have another illustration of the law that *the current developed by induction flows in such a direction as to oppose the motion which produces it.*

If a conductor is now moved to the right, that is, away from the magnet with its accompanying field, Fig. 315, the lines of force are wrapped in the opposite direction about it, and a current is generated flowing in the opposite direction, in this case away from the observer. Here also, by the repulsion of the lines of force on the side of the conductor which is in advance, it is evident that the current generated flows in such a direction as to oppose the motion.

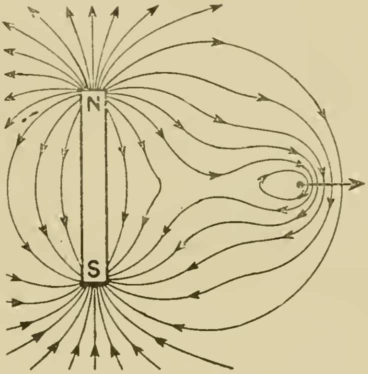

FIG. 315. Current in the opposite direction induced by carrying a current away from a magnet.

An extension of this same reasoning shows that if the pole of a magnet is thrust into a solenoid in which no current is flowing, the lines of force of the magnet will wrap themselves about the conductors, thus generating a current in the solenoid. Moreover, in order that work may be done, the induced current thus generated will flow in such a direction that *the pole of the solenoid into which the magnet is carried will be of the same nature as the pole of the magnet itself,* so that these two would repel each other, and thus tend to oppose the motion whereby the current is generated.

Thus, if the north pole of the magnet is thrust into the solenoid, the elastic, rubber-like lines of force are caught and retarded by the conductors of the solenoid, Fig. 316, so that they are bent back upon themselves, and therefore emerge from the end of the solenoid into which the magnet is carried. These lines of force which are thus wrapped about the conductors and which emerge from the pole of the solenoid, *are the lines of force of the current which is thereby generated.* But the pole of a

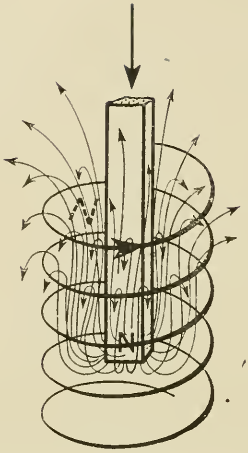

FIG. 316. The north pole of a magnet thrust into a solenoid.

solenoid from which lines of force emerge is a north pole; and hence the current generated flows in such a direction as to develop *a north pole at the end of the solenoid into which the north pole of the magnet is carried.* Since these two north poles repel each other, and since the current is generated only by bringing the two together, we see that *the current flows in such a direction as to oppose the motion which produces it,* in accordance with Lenz's law.

Inasmuch as the induced current flows only as long as the motion continues, the flow soon ceases, and the field of force about the magnet assumes its normal condition, the lines of force then passing between the conductors of the solenoid and even through them. Accordingly, when the magnet is withdrawn from the solenoid, the lines of force are again wrapped about the conductors, this time, however, in the opposite direction, so that a current is generated in the opposite direction, and a pole of the solenoid is created of a nature opposite to that of the magnet which is being withdrawn. In this case, the lines of force, which are caught and therefore retarded by the conductors, Fig. 317, are enormously stretched as the magnet is carried away, and for a short time still enter the end of the solenoid from which the magnet is being removed. But the pole of a solenoid which lines of force enter is a south pole, so that *a south pole is developed at the end of a solenoid from which the north pole of a magnet is being withdrawn.* Since these opposite poles attract each other, thereby tending to prevent their separation, and since

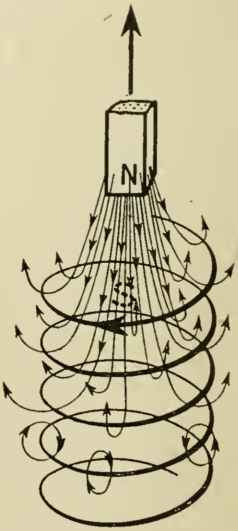

FIG. 317. The north pole of a solenoid withdrawn from a solenoid.

the current in this direction is generated only by this separation, we see again that *the induced current flows in a direction such as to oppose the motion which produces it,* and Lenz's law is obeyed.

The magnetic field of force about a conductor carrying a current is likewise able to induce a current in a neighboring parallel conductor when the two are moved with respect to each other. Let us suppose that the wire *P*, in which the current is flowing away from the observer, Fig. 318, is suddenly brought up to the parallel conductor *S* which does not contain a current. As the field of force of *P* approaches the conductor *S*, the lines of force of the field of *P* wrap themselves in such a way about *S* that a curent is induced in *S* flowing towards the observer. Thus, as the two are brought together, the induced current flows in a direction opposite to the inducing current, and this would tend to make the two conductors repel each other (p. 553) and move apart. But as the induced current is generated in this direction only by bringing the two together, we see that the induced current flows in such a direction as to oppose the motion which produces it.

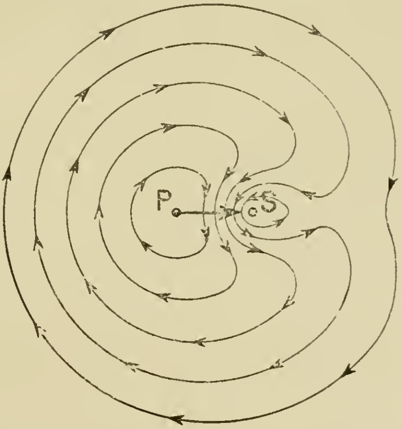

FIG. 318. An inverse current is induced in *S* by the approach of P.

The induced current, however, last only as long as the motion continues, so that when this ceases, the lines of force of *P* pass completely around *S*, as if the latter were not present. Therefore, when *P* is suddenly removed from *S*, that is, carried to the left, Fig. 319, the lines of force, which are due to *P*, and which also pass around *S*, are distended, and wrap themselves about *S* in the same direction as those about *P*. This induces a current in *S* flowing in the same direction as that in *P*, or away from the observer.

In practice, the conductor *P* in which the current is flowing is called the *primary* circuit, and the conductor *S* in which the current is generated by induction is called the *secondary* circuit. We see, then, in general, that the *approach* of the primary circuit generates a current in the secondary flowing in a direction *opposite* to that in the primary, or an *inverse* current, while the *removal* of the primary circuit generates a current in the secondary flowing in the *same* direction as that in the primary, or a *direct* current.

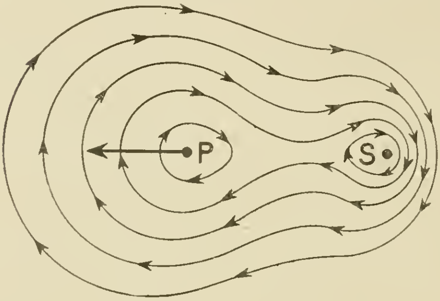

FIG. 319. A direct current is induced in S by the removal of P.

An examination of the last two figures will enable us to deduce another fact regarding induction. Let us assume that the primary and the secondary conductors are lying parallel to each other, but with no current in either. If, now, a current is sent through the primary, the circles of the lines of force suddenly sweep out from this conductor, thus forming the field of force about that conductor. As they suddenly grow out from this conductor as a center, they encounter the secondary conductor, about which they wrap themselves as they do when the primary is brought towards the secondary. Therefore *closing the circuit through the primary develops a current flowing in the opposite direction in the secondary.* Let us now imagine the current to have become steady in the primary, and the field of force of the primary to pass around the secondary, the temporary induced current in the latter having ceased. If the current in the primary is then *broken*, the circular lines of force sweep in toward this conductor, tending to contract about it as a center. As they meet the secondary, they are distended in their attempt to cut across this conductor, and so wrap themselves about the latter in the *same* direction as the lines of force about the primary. Accordingly, *breaking the current in the primary produces the same effect as separating the two circuits,* *that is, it induces a current in the secondary flowing in the same direction as the current in the primary.*

The phenomena which are here described are also produced when the primary and secondary conductors are wound into solenoids. If a current is flowing in the primary solenoid, and if this is *approached* to the secondary, Fig. 320, a current is developed in the latter in the *opposite* direction to that in the primary. In other words, a current is developed in the secondary in such a way that the two solenoids have like poles towards each other, and then, by their mutual action they repel each other. If now, the primary is *removed* from the secondary, Fig. 321, a current is induced in the latter flowing in the *same* direction as that in the primary. That is, the current in the secondary flows in such a way that this solenoid develops a pole of a nature opposite to that of the primary, and these, by their mutual attraction tend to prevent the separation.

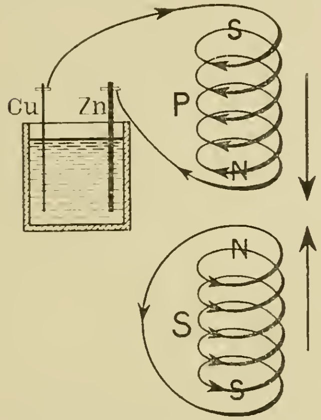

FIG. 320. The approach of *P* induces an inverse current in *S*.

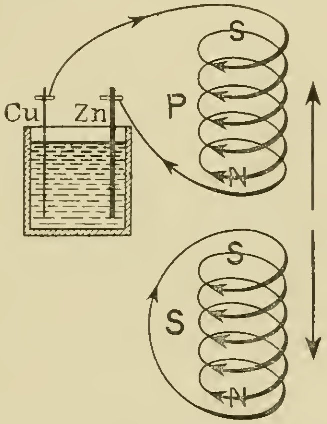

FIG. 321. The removal of *P* induces a direct current in *S*.

Here also, as before, currents may be induced in the secondary simply by making and breaking the circuit in the primary. If the two solenoids are placed together, and if the current is *made* in the primary, the field of force which emanates from this solenoid sweeps out over the conductors of the secondary, producing a current in the latter in a direction *opposite* to that in the primary. When the circuit in the primary is *broken*, the lines of force, in contracting toward this circuit, move across the conductors of the secondary precisely they would if the primary were removed. Therefore, *bringing the primary towards the secondary, or making the circuit in the primary,* induces a current in the secondary in the *direction opposite to that in the primary, or an inverse current.* In this case the two currents flowing in opposite directions repel each other that is, the current developed in the secondary flows in such a direction as to oppose the approach of the two coils, to which its existence is due. *Removing the primary from the secondary, or breaking the circuit in the primary, produces a current in the same direction as that in the primary, or a direct current.* Here the two currents, being in the same direction, attract each other and thus tend to prevent the separation which gives rise to the induced current. Thus, in every instance of the production of an induced current, Lenz's law applies.

It is to be observed in this connection that, as the primary coil is *approached* to the secondary, *the number of lines of force passing through the secondary is increased,* and the nearer together the two are brought the greater is the number of lines of force which pass through the secondary. Moreover, as the primary is *withdrawn* from the secondary, *the number of lines of force passing through the latter is decreased.* This serves to illustrate a law which is of general application, and which can often be employed where a direct cutting of the lines of force is not obvious, namely, *an increase in the number of lines of force passing through a circuit produces an inverse current, while a decrease in the number of lines of force passing through a circuit produces a direct current. The induced current, therefore always flows in such a direction as to maintain unchanged the number of lines of force passing through the area enclosed by its circuit.*

497. The Transformer. The principles which have here been discussed are those which are at the basis of the action of all induction coils, or transformers. If a bar of soft iron, or better, a bundle of soft iron wires is placed in the primary solenoid, the effect upon the secondary will be greatly increased, due to the fact that the lines of force of the iron core, which has now

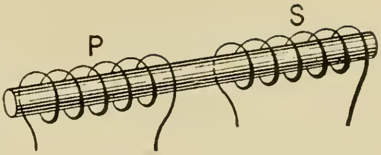

FIG. 322. Primary and secondary wound upon an iron bar.

become magnetic, are added to the lines of force of the primary solenoid. Indeed, if this iron core is extended, Fig. 322, and the secondary wound upon the remote end, making and breaking the circuit in the primary produces inverse and direct currents in the secondary of nearly the same strength as those in the primary, provided the coils are similar in every way. If the two ends of this iron are now brought together so as to form a ring, upon which the primary and secondary coils are wound entirely separated from each other, Fig. 323, we have the essentials of the transformer, which is so generally used for electric lighting at the present time. When the circuit is closed in the primary, an inverse current is at once developed in the secondary, and when the current in the primary is broken, a direct current is generated in the secondary. If, therefore, as is the case in practice, the primary is supplied with an alternating current, that is, one which reverses its direction a great many times a second, a current is developed in the secondary which changes its direction as often as does that of the primary.

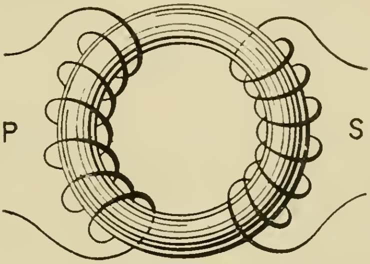

FIG. 323. The transformer.

If the transformer is reversed and a current is sent through the secondary, a current will then be induced in the primary, so that either coil may be used as the primary, inasmuch as the induction is mutual. Moreover, if the iron ring is made of soft iron, the difference of potential developed at the two ends of the secondary bears the same relation to the difference of potential used on the primary as the number of turns of wire on the secondary solenoid bears to the number of turns on the primary. Therefore, if the coils are wound with different numbers of turns of wires, the difference of potential developed at the ends of the secondary will be greater or less than that used on the primary according as the secondary has a larger or a smaller number of turns. In this way we can have a "step up" transformer if the secondary has a greater number of turns than the primary, or a "step down" transformer if the secondary has a smaller number of turns than the primary.

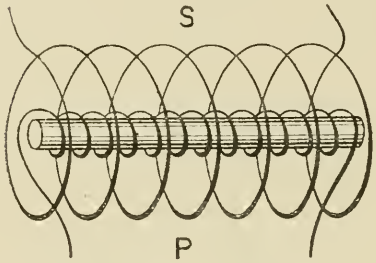

FIG. 324. The induction coil.

In the induction coil, Fig. 324, the primary circuit consists of a comparatively few turns of heavy wire wound upon a core composed of a bundle of soft iron wires, while the secondary consists of many thousand turns of very fine wire wound wholly separated and perfectly insulated from the primary. Since the number of turns in the secondary is perhaps many thousand times as great as the number of turns on the primary, the difference of potential generated by a large coil is enormous, and rivals that developed by static means.

However great may be the ratio of transformation, the energy of the current remains constant throughout, except for certain losses, such as hysteresis etc. Thus, if the current and the electromotive force of the primary are represented by I and E, while the current and the electromotive force of the secondary are represented by I' and E', it follows at once from the principle of the conservation of energy, that

$$IE = I'E'$$

Since E' is many thousand times greater than E in the ordinary induction coil, it follows that the current which flows in the secondary of such a coil is very small indeed, although the potential which is produced by this means is exceedingly high.

498. Foucault Currents. We have seen that whenever a conductor and a field of force are moved with reference to each other, a current is developed in the conductor, and this, by its reaction upon the field of force, flows in such a direction as to oppose the motion which produces it. This is equally true when large metallic masses are moved in a magnetic field, or when the intensity of a field of force is made to vary in the neighborhood of large masses of metals. In both of these cases currents are generated in the metallic masses, and these currents flow in closed circuits through the metal, forming little eddy currents within the mass of the metal itself.

The generation of these eddy currents may be beautifully illustrated by swinging a plate of copper between the poles of a powerful electromagnet. In Fig. 325, if we imagine the north pole of the magnet to be on the side of the copper towards us, and the south pole to be on the farther side, we shall be looking in the direction of the lines of force of the magnet as they stream through the copper plate from the north pole to the south. If the copper plate is now allowed to swing to the right and left across these lines of force, currents will be generated which will flow in little eddies, or whirls, forming closed circuits within the mass of the copper. Since these currents by Lenz's law, flow in such a direction as to oppose the motion which produces them, the plate will quickly be brought to rest.

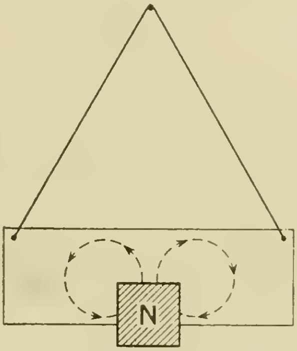

FIG. 325. Generation of Foucault currents in a swinging copper plate.

Eddy currents which are thus generated in a mass of metal when this is moved in a magnetic field are known as *Foucault currents.* Energy is wasted if these currents are allowed to form, since they cannot in any way be utilized, and, indeed, they tend only to impede the motion of the metal in which they are generated. The best way of avoiding their formation is to use thin plates of metal instead of heavy masses, and to separate these plates from one another by paper, which prevents these Foucault currents from flowing. In the above illustration, if the copper plate is sawed into strips, Fig. 326, these eddy currents cannot flow, and the plate, which before was immediately arrested, now continues to move almost as freely as if it were not swinging in a magnetic field.

In the construction of the armature of the modern dynamo, the conductors are wound upon a heavy iron core. Inasmuch as the current in these conductors continually varies in intensity and even changes in direction, the field of force due to these conductors is a continually varying one. This constantly varying field of force, occurring in the immediate neighborhood of the iron core, would cause eddy currents to be formed if this were a solid mass of iron. To obviate this difficulty, the iron core is built up of thin plates, or laminae, separated from one another by paper, so that the Foucault currents, which would eddy through the iron parallel to the conductors, are prevented from forming.

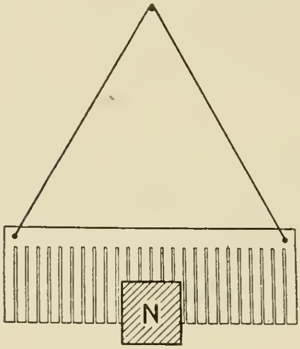

Fig. 326. Foucault currents prevented from flowing by sawing the copper plate into strips.

499. The Principle of the Dynamo. The modern dynamo is a device for producing an intense magnetic field of force, and for carrying a series of conductors across the lines of force of this field, thereby generating a current in these conductors. In the early machines the magnetic field was produced by large permanent magnets, but these are now replaced by the much more powerful electromagnets.

To understand the action of one form of dynamo now in common use, let *N* and *S*, Fig. 327, represent the north and south poles respectively of the electromagnet, hollowed out somewhat to receive the rotating portion, which is called the armature. The two arms of this electromag-

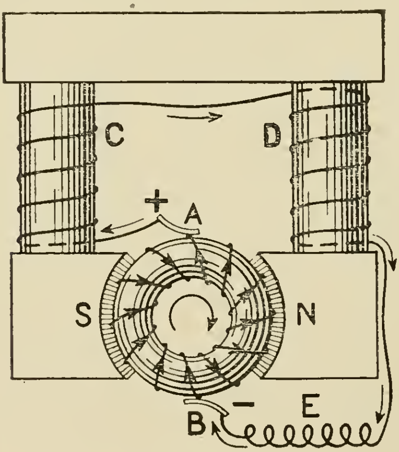

Fig. 327. The series dynamo.

net, to which the intense magnetic field is due, are called the field magnets. The armature is made in many forms, and of these perhaps the simplest consists of a ring made of laminated iron, around which is wound a coil of copper wire. This wire is wound on the ring in one continuous coil, being carried over the outside and through the center, and so on clear around the ring, until finally the two ends are brought together and soldered. This is the conductor which is to be used for cutting across the lines of force of the magnetic field, and in this the current is to be generated by induction. To this end, the iron ring, with the endless conductor wrapped around it, is placed in the space left for this purpose between the north and the south poles of the field magnets, and there it is made to rotate upon its axis, being driven by a steam engine or other source of power.

Since the soft iron of which the ring is made has a very high permeability, the lines of force of the field stream from the north pole of the magnet across the small air gap into the iron ring, thence through the iron to the opposite side of the ring, then again across the small air gap into the south pole of the magnet. The circuit of the lines of force is then completed through the arms of the electromagnet back again to the north pole. It is to be observed that this entire circuit of the lines of force is wholly in iron except the small air spaces which are necessarily left between the north pole of the magnet and the iron core of the armature on the one side, and between the iron core of the armature and the south pole of the magnet on the other side. Although these "air gaps" are made very small, it is here that the endless conductor which is wound upon the armature cuts across the lines of force of the field, thus generating the current. It is, moreover, to be borne in mind that it is only the turns of wire which lie on the outside of the ring that are effective in cutting the lines of force, inasmuch as it is only wires on the outside of the ring that encounter the lines of force of the field as they flow from the north pole into the ring, and from the ring across the air gap into the south pole of the magnet. The wire, therefore, that is wound upon the inside of the ring is quite useless so far as assisting in the generation of a current, and, indeed, it is worse than useless, inasmuch as the current which is generated must flow through it, thereby introducing resistance.

As the armature now turns, the conductors upon its exterior cut across the lines of force in one direction when near the north pole, and in the opposite direction when near the south pole. This will therefore develop currents flowing in one direction in one half of the armature and in the opposite direction in the other half. Assuming that the armature turns in the direction of the arrow, a given conductor on the outside of the ring cuts lines of force continuously in one direction during the half revolution from *A* to *B*, and in the opposite direction during the half revolution from *B* to *A*. Currents are thus generated flowing in opposite directions in the two halves of the ring at the same time. That is, if the current flows toward the point *A* on the side of the armature near the north pole, it will also flow toward the point *A* on the side of the ring near the south pole. Accordingly, from the point *B* the current flows through both halves of the armature toward the point *A*, thus making *A* a point of high potential, and leaving *B* a point of low potential. If a spring or metallic brush is made continuously to touch the point *A*, this brush will be maintained at high potential, and in the same way a brush touching the point *B* will always be at low potential. If, therefore, the points *A* and *B* are connected by a wire through any circuit, a current will flow through this circuit, and will be maintained as long as the process of turning the armature is continued.

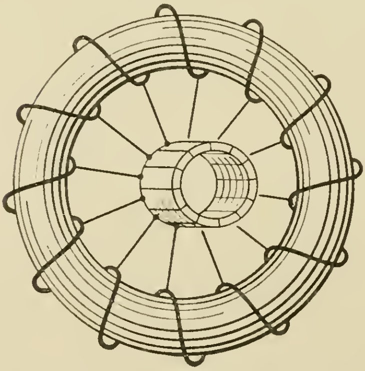

FIG. 328. The commutator, showing its connection with the armature.

In practice it is not usual to have the springs or brushes touch the outside of the armature, as here represented, but it is customary to make use of a commutator, Fig. 328, which enables contact nevertheless to be established with the conductors of the armature. This commutator consists of a series of copper bars, insulated from one another and collected in the form of a cylinder, each bar being connected by a wire with one of the turns of wire on the armature. These copper bars, therefore, are at the same potential as the coils of the armature with which they connect, and it is upon these that the brushes press and thus form contact with the wires.

The current which flows from the armature, or at least a portion of it, is now carried through a conductor wound around the arms of the field magnets, thereby using the very current which the machine generates for the purpose of magnetizing, or exciting, the fields. According to the particular method in which this current is carried around the field magnets, the dynamo is either a series, a shunt, or a compound machine.

500. The Series Dynamo. In the series machine, the entire current which the armature generates is carried around the arms of the electromagnet, as at *C* and *D*, Fig. 327. After passing from *A* around the field magnets, the current then flows out through the external circuit, as at *E*, where it is used for lighting lamps, driving motors, etc. It then flows back to the low potential brush at *B*, thus completing the circuit, having traversed the field magnets and the external resistance *in series*. When the series dynamo is started, the conductors upon the outside of the armature cut across a very weak field of force, inasmuch as the only magnetism now existing is the small amount of residual magnetism remaining in the poles of the magnets. This generates a feeble current which flows around the field magnets, thus making them more strongly magnetic, and therefore increasing the strength of the field. The stronger field enables a stronger current now to be generated, and the increased current makes the field magnets stronger and the magnetic field still more intense.

This action between the field strength and the current, therefore, is a reciprocal one, the increase in the current bringing about a stronger magnetic field, and the more intense magnetic field, in turn, serving to increase the strength of the current. This cumulative process, however, cannot continue indefinitely, inasmuch as the field magnets become saturated (p. 577), and this state of the iron places a limit to the strength of the magnetic field that may be obtained. The field magnets, therefore, are wound with such a number of turns of wire that a current of a certain strength will produce saturation, and this current is the normal current which the machine is designed to generate. A series dynamo, therefore, is a constant current machine.

501. The Shunt Dynamo. In the shunt dynamo, Fig 329, the current from the armature divides into two parts, one portion passing through many turns of fine wire around the field magnets, while the remainder is used for the external circuit. Since the resistance of the circuit around the field magnets is constant, a nearly constant current flows through this shunt circuit, so that the strength of the magnetic field of the machine remains practically always the same. Accordingly, if the armature rotates with uniform speed, the rate at which lines of force are cut is nearly constant, and the electromotive force of this machine is therefore also nearly constant. While a series dynamo is serviceable for running arc lamps, where a constant current and a high electromotive force are necessary, a shunt machine is used for incandescent lighting, where a constant difference of potential must be maintained.

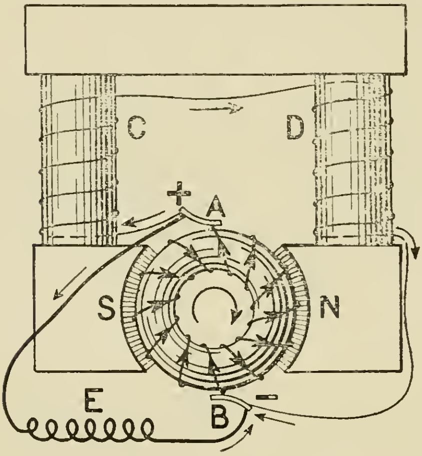

FIG. 329. The shunt dynamo.

The electromotive force of the shunt dynamo, while nearly constant under ordinary conditions, is not rigidly so, and may, indeed, at times fluctuate between wide limits. The difference of potential produced by this machine is highest when the external circuit is open, and when the full electromotive force of the machine sends the current through the field magnets. When the external circuit is closed, the flow of current through this circuit lessens the difference of potential between the brushes, and, accordingly, not so much current flows through the shunt circuit, so that the electromotive force of the machine falls. Indeed, if this dynamo is short-circuited, that is, if the brushes are connected by a short piece of heavy copper wire as the external circuit, the difference of potential between the brushes falls almost to nothing. This decreases the difference of potential between the two ends of the field circuit to an equally small amount, so that practically no current then flows around the field magnets, and the excitation of the machine is destroyed. A shunt dynamo, therefore, will "run down" if short-circuited, so that a machine of this type can never be injured by drawing too large a current. In order to maintain a constant difference of potential with a shunt machine, a resistance is placed in the field circuit, and this resistance may be decreased by hand as the electromotive force of the machine begins to fall, thereby sending a stronger current around the fields and thus keeping up the required constant electromotive force.

502. The Compound Dynamo. In the compound dynamo, Fig. 330, the principles of the series and the shunt machines are combined, so that the regulation for constant electromotive force may take the place automatically. In the series machine, the magnetic field increases as the current increases, while in the shunt machine the excitation decreases as the current in the external circuit becomes greater. To maintain a constant electromotive force, therefore, a few turns of the wire forming the external circuit are wound as series coils around the field magnets of what would otherwise be a shunt machine. Moreover, the turns of this series circuit are so adjusted in number that the increase in the electromotive force which they pro-

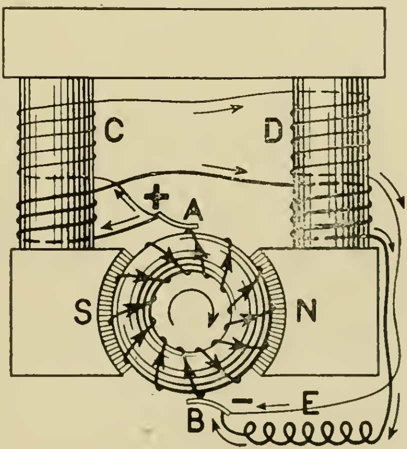

Fig. 330. The compound dynamo.

duce as the current increases just compensates for the decrease in the electromotive force which occurs in the shunt machine, so that for whatever current the machine produces, the electromotive force remains constant. In the modern compound machine, this end is very nearly completely attained.

503. The Alternating Current Dynamo. From a study of the armature of the dynamo, we have seen that the current at any instant flows through opposite halves of the armature in opposite directions; that is, the current in any particular coil flows one way for half a revolution, and in the opposite direction for the other half revolution. Accordingly, if the armature consists of a single coil of wire, Fig. 331, and if the two ends of this coil are connected to rings concentric with the axis upon which the armature turns, permanent connection may be made with the ends of this coil by allowing brushes to press upon these "collector rings." A current flowing first in one direction and then in the other —that is, an alternating current—will be generated in this coil of wire, and an alternating current will flow through the external circuit which connects the brushes.

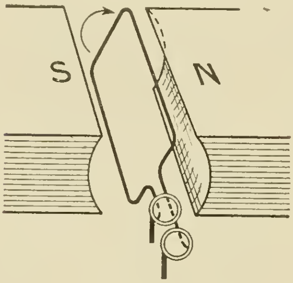

FIG. 331. The generation of an alternating current.

In practice, however, it is usual to design the alternating current dynamo something in the following way: The field magnets consist of a number of pairs of north poles and south poles mounted upon the inside of a large iron ring and pointing towards the center, Fig. 332. These are magnetized by winding wire in opposite directions around these poles and sending a direct current from an outside source through this wire. The armature consists of as many coils of wire as there are magnetic poles in the field, all mounted upon the circumference of a rotating drum so as to pass very near the pole pieces. Adjacent coils in this armature are wound in opposite direction so that the current generated in one coil, as it passes before a north pole, flows in the same direction as that induced in the next coil as it passes before a south pole. The current thus generated in the armature changes direction, of course, when the drum turns so that the coils which were opposite one set of poles now pass before those adjacent, which are of opposite nature. As the drum turns farther, the current generated in these coils again reverses in direction, and this reversal in direction takes place as often as each coil leaves one magnetic pole, and approaches another of opposite polarity.

It is thus seen that there are as many reversals of the current during one revolution of the armature as there are magnetic poles in the field. A pair of reversals, that is, one complete reversal each way, is known as a *cycle*, from which it follows that there are as many cycles during one revolution as there are pairs of poles constituting the field magnets. The time occupied for completing one cycle is known as the *period*, and the number of cycles per second is called the *frequency*. The two ends of the armature wire are connected to collector rings, and these are always in contact with the metallic brushes which press upon them. Connection may be made with these metallic brushes, so that the alternating current generated in the armature may flow away to the external circuit.

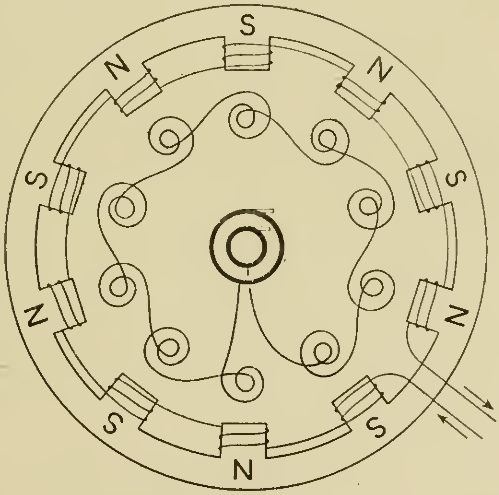

FIG. 332. Diagram of the alternating current dynamo.

504. Economy of Transmission of Alternating Currents. The general use of the electric light througout the civilized parts of the world renders it necessary often to carry currents long distances from the central station where it is generated. When this is attempted with a direct current, it is found that the unavoidable resistance of the conducting wire wastes so much of the energy of the current as heat, that it is impossible to carry the current to any great distance without employing enormous conductors in order to decrease the resistance. Inasmuch as the power of the current is the product of the current and the electromotive force (p. 519), it is evident that if the electromotive force were made very great, the strength of the current could be made correspondingly small, and the amount of energy wasted as heat could thus be diminished. Since it is unsafe to bring a current of extremely high voltage into dwelling houses, a limit for the electromotive force of such a circuit is set at about 110 volts.

With an alternating current, however, it is possible to use a "step up" transformer at the station, and thus send a large amount of electrical energy over long distances as a small number of amperes, but at extremely high voltage, and then to employ a "step down" transformer at the house where the current is to be used, and thus obtain again a large current at a comparatively low electromotive force. During the transmission between the station and the house, therefore, the number of amperes of current is so small that the amount of energy which is wasted as heat does not offer a serious difficulty, even when a relatively small wire is used as the conductor. Accordingly, the alternating current is employed almost universally for electric lighting, and is now being used also for the transmission of power over long distances.

505. The Electric Motor. Whenever an electric current from an outside source is sent through a conductor in a magnetic field, a mutual reaction takes place between the lines of force due to the current and those due to the field. From the fact that the lines of force of the field are in general straight, while those due to the current are circular, the lines of force of the field will be in the same direction as the lines of force due to the current on one side of the conductor, and in a direction opposite to that of the lines of force due to the current on the other side of the conductor. On the side, therefore, where the lines of force of the field are in the same direction as those due to the current, repulsion between these two takes place, while on the other side, where the lines of force from these two sources are in opposite directions, the circular lines of force due to the current tend to open out and join with those of the field, and attraction occurs. If the conductor, therefore, is free, it will sweep across the lines of force of the field, *always moving away from the side on which the lines of force of the field are in the same direction as those about the current.*

Thus, Fig. 333, let the black dots represent the cross-section of a conductor through which a current from an outside source is flowing away from the observer in the earth's uniform field. On the right hand side of the conductor, the lines of force due to the current are in the same direction as those of the field, hence on this side repulsion between these two takes place. On the left hand side of the conductor, the lines of force of the field and those due to the current are in opposite directions, hence these tend to attract each other. The lines of force due to the current, therefore, are repelled on the right side and attracted on the left hence the conductor in this case moves to the left.

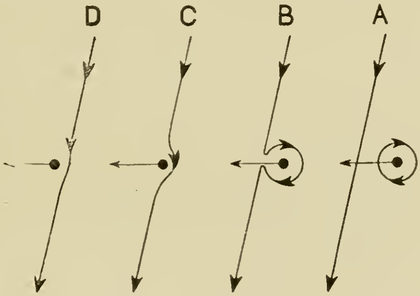

FIG. 333. The principle of the electric motor.

The generation of a current by cutting across the lines of force of a magnetic field is the principle of the dynamo. The production of motion by sending a current through a conductor in a magnetic field, thereby effecting a reaction between the lines of force of the field and those of the current, is the principle of the electric motor. Any device, therefore which can be used as a dynamo for generating an electric current may also be used as a motor by sending current into it from an outside source. Accordingly, an electric motor is identical in form and construction with a dynamo, and may in every instance be used as a dynamo.

It is carefully to be observed, however, that if a current is *generated* in the moving conductor by carrying it in a certain direction, *the conductor will move in the opposite direction when a current is sent into it from an outside source,* the direction of the current in the conductor remaining the same in both instances. Thus, we have seen, Fig. 308, that a current is generated flowing away from the observer by carrying the conductor to the *right* across the earth's field, while here we see that if we send a current through the conductor flowing away from the observer, the conductor will move to the *left*. So also, in the series dynamo, (p. 604), we saw that by turning the armature clockwise a current is generated flowing through both halves of the armature in multiple from *B* to *A*. If the dynamo is now stopped and a current from an outside source is sent through the same circuit flowing from *B* to *A*, the armature will turn counter-clockwise. The armature of the machine when used as a motor, therefore, will turn in the opposite direction from that in which it is turned when used as a dynamo, the direction of the current through the machine being the same in both cases.

In a shunt machine, the direction of rotation as a motor is the same as its direction of rotation as a dynamo, since, when it is used as a motor, the field is reversed. This reversal of the field, together with its reversed direction of motion as a motor, causes the direction of rotation of the shunt motor to be the same as that of the shunt dynamo, the current through the armature being the same in each case.

506. Reaction Between an Electric Current and a Magnetic Field. A most convenient and instructive device for studying the mutual action between a movable current and a magnetic field may be obtained by suspending a rectangular frame carrying a current so that it will swing freely before a magnet which may be brought in various positions near it, Fig. 334. In each case the conductor, if free, will move across the lines of force of the magnetic field, and *always away from that side where the lines of force due to the current are in the same direction as the lines of force of the field.* It will then be found that the direction in which the current moves as a motor, that is, the direction in which a current always tends to move when placed in a magnetic field, may be determined by the following modification of Ampere's rule: *Swim in the current and with the current, facing in the direction of the lines of force of the field, then the current tends to move to the left and the lines of force to the right.*

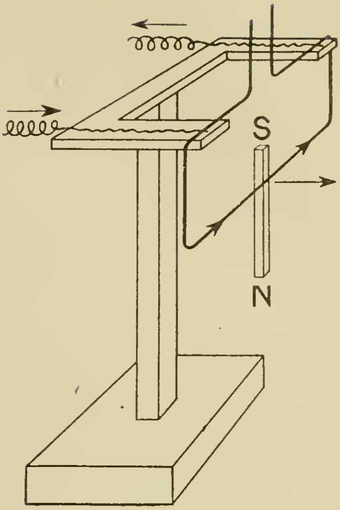

FIG. 334. The movement of a conductor in a magnetic field.

This rule applies to all cases in which the motion of a conductor is effected by causing a current to flow through that conductor in a magnetic field. If we refer to Fig. 313, illustrating Faraday's disc dynamo, we see that by turning the disc in the direction of the arrow, a current is generated flowing from the axis to the circumference of the disc. If now, from an outside source, a current is sent through this disc flowing from the axis to the circumference, the disc will turn in the opposite direction. Possibly the simplest illustration of the action of the electric motor is obtained by the experiment where two copper tanks containing an electrolyte are each provided with a little copper boat, Fig. 335, these two copper boats being connected by a stout copper wire. The wire uniting the boats forms a movable conductor in the earth's magnetic field, and hence this conductor moves across the lines of force of the earth's field whenever a current is sent through it. Moreover, the direction of motion is always away from that side of the conductor where the lines of force due to the current are in the same direction as the lines of force of the field.

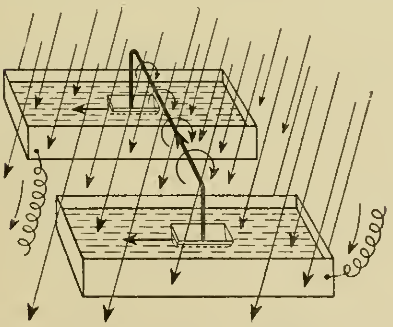

FIG. 335. A conductor moving in the earth's magnetic field.

507. The Counter Electromotive Force Generated by an Electric Motor. Whenever the conductors of an electric motor move in a magnetic field, even though they move because of the reaction between the lines of force due to the current and those of the field, these conductors cut across the lines of force of the field. Moreover, they move in such a way as to tend to develop a current flowing in a direction opposite to that of the current already flowing. Thus, in Fig. 308 we have seen that a current flowing away from the observer is developed by carrying the conductor to the right. By the principle of the motor, however, we see that if we send through this conductor a current from an outside source flowing away from the observer, the conductor will move to the left. But from Fig. 309 we have seen that if the conductor moves to the left across the same field of force, a current will be generated flowing towards the observer. Therefore, if the conductor carries a current from an outside source flowing away from the observer, and this conductor moves to the left, a counter current will be generated, due to this motion, tending to flow toward the observer. That is, the counter current which is thus generated will flow in a direction opposite to that of the current which is already flowing. If, therefore, this conductor is regarded as one of the wires of an electric motor and carries a current flowing away from the observer, the conductor, in moving to the left, will generate a counter current, due to its motion, tending to flow towards the observer. Accordingly, whenever a motor runs, *a counter current is developed flowing opposite to the current which is sent into the motor from the outside source.* This is usually expressed by saying that *the motor develops a counter electromotive force, acting in a direction opposite to that of the electromotive force of the current which drives the motor.*

Indeed, this principle of the generation of a counter electromotive force has already been set forth (p. 584), where the rotation of a semicircular current about a unit magnetic pole was discussed. It was there shown that if E is the difference of potential between the points A and B when the conductor is at rest, the resultant electromotive force between the same points is only $E—e$ when the conductor is in motion. Here e is the electromotive force generated by the movement of the conductor across the lines of force of the magnetic field, and represents *the rate of cutting these lines of force.* Furthermore, it is seen that this electromotive force which is thus generated *acts in a direction opposite to that of the electromotive force already existing,* and therefore is a counter electromotive force.

Accordingly, if E denotes the electromotive force of the outside source which sends current into a motor, and R denotes the resistance of the motor, then the current C which flows through the motor when the armature is prevented from turning is

$$C=\frac{E}{R}\ .$$

If, however, the armature of the motor is allowed to turn, this, as we have seen, develops a counter electromotive force equal to e and acting in a direction opposite to E. Therefore, the resultant electromotive force which is effective in sending current through the machine when the motor is running is only

$$E-e,$$

and consequently the current which flows through the motor when the latter is running is

$$C=\frac{E-e}{R}\ .$$

e depends, however, upon *the rate of cutting the lines of force* (p. 584), and therefore upon the *speed* of the motor. That is, the faster the motor turns, the greater becomes e, and therefore the smaller becomes C. The faster the motor runs, therefore, the less is the amount of the current which is required to run it. Moreover since the counter electromotive force e is developed by doing work outside of the electric circuit, or external work, it follows that the larger e becomes, the larger, also, becomes the portion of the energy of the current which is converted into useful external work. It is on this account that it is more efficient to run all electric motors at a very high speed.

508. Illustration of the Counter Electromotive Force. The generation of a counter electromotive force when the motor runs may be nicely illustrated by placing a little incandescent lamp in series with a motor, so that the current which flows through the motor flows also through the lamp. When the current from the outside source is sent through the motor as this is standing still, the lamp lights up brightly, showing that a large current is flowing. As soon, however, as the armature begins to turn, the lamp becomes dim, showing that not so much current is flowing, and therefore that the counter electromotive force is opposing the direct electromotive force. Finally, when the armature of the motor has attained a high speed of rotation, the lamp does not light up at all, showing that the current passing through the motor, and therefore through the lamp, is not enough even to heat the lamp red hot. This illustrates how enormously the current through the motor decreases as the speed of the armature increases.

As a matter of fact, this experiment could not be performed as here described, since the little incandescent lamp could not carry sufficient current to run the motor. A resistance, therefore, is put in series with the motor, Fig. 336, and the little lamp is put in multiple with this resistance. The little lamp and the resistance together carry the current which passes through the motor, and of this total amount of current the resistance carries about 9/10 and the lamp 1/10. The amount through the lamp, therefore, is always proportional to the total amount of current flowing through the motor, so that when the lamp lights up we know that a large current is flowing through the motor, and when the little lamp falls in brightness or fails to light up altogether, as it does when the armature attains a high speed, we know that the current through the motor is greatly diminished.

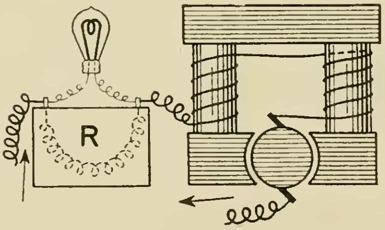

FIG. 336. Experiment showing the counter electromotive force of an electric motor.

509. Counter Electromotive Force of Self-Induction. Whenever a current is sent through a solenoid consisting of many turns of wire, lines of force are suddenly introduced into the core of the solenoid, precisely as if a magnet had been thrust into the opening at the center. But this, we know (p. 594), develops a current flowing in such a direction as to oppose the introduction of the magnet. Accordingly, when a current is started in a solenoid, an electromotive force is developed acting in a direction opposite to that which causes the current to flow. When the curent in the solenoid is interrupted, the lines of force disappear from the center, just as if a magnet had been withdrawn from the solenoid. But this, as we have learned, develops a current in the solenoid tending to prevent the removal of the magnet. It follows, therefore, that when a current in a solenoid is broken an electromotive force is thereby developed acting in the same direction as that which already exists. Accordingly, when a current is *closed* in a solenoid, an *inverse* electromotive force is thereby generated, and when a current is *broken* in a solenoid, a *direct* electromotive force is created.

From this it may be seen that each turn of wire in a solenoid in which a current is increasing or decreasing acts as a primary coil with respect to every other turn which acts as a secondary. There is thus produced the phenomenon of *self-induction*, whereby a current which is increasing or decreasing in a solenoid acts inductively upon the very circuit in which it itself is flowing. Moreover, it will be noticed that the *direction* of the electromotive force produced by self-induction is always such as *to oppose the current if this is increasing, and to assist the current if this is decreasing.* In other words, the self-induction acts in such a way as *to resist any change in the current which flows through the solenoid, and therefore to oppose any change in the number of lines of force which pass through the circuit.* That is, if no current is flowing in a solenoid, the self-induction of this coil will be in such a direction as to oppose the current when one is formed, and if a steady current is flowing in a solenoid, the self-induction will tend to maintain this current unchanged as soon as it is broken. *The effects of self-induction, therefore, are always in such a direction as to prevent any change in the intensity of the current flowing.* In this respect the phenomena of self-induction, which tends to maintain the existing condition as regards current, are analogous to those of inertia, which tends to maintain unchanged the existing condition of matter as regards motion. Self-induction, therefore, presents phenomena analogous to an electromagnetic inertia.

Owing to the fact that self-induction always opposes any change in the strength of the current, a current which is sent into an inductive circuit rises to the maximum value only slowly and after the lapse of an appreciable time, and in the same way a current that is decreasing in such a circuit is delayed in this change also through an appreciable time. Accordingly, if an alternating current is sent through a solenoid, particularly if the latter contains an iron core, the alternations of the current take place in far shorter time than is required for the current to attain its full value in the solenoid, and therefore the current which actually flows is but a small part of what would flow through a resistance equal to that of the solenoid if the current were direct and steady. A solenoid containing iron is therefore known as a "choking coil," because it thus reduces the amount of current which would flow through a non-inductive resistance equal to that which the coil contains.

Accordingly, the action of self-induction for an alternating current is analogous to that of resistance for a direct current, in that both tend to reduce the strength of the current which flows. One important distinction between the two, however, must constantly be borne in mind. When resistance is introduced to diminish the intensity of a direct current, the energy represented by the current flowing through this resistance is wasted as regards useful work, and appears only as heat. But when the self-induction of a choking coil is used to cut down an alternating current, no energy is lost, since the excessive current simply does not flow, inasmuch as this is prevented from flowing by the self-induction of the coil.

A most instructive illustration of the effect of self-induction in a circuit which contains iron may be obtained by connecting an incandescent lamp in parallel with a large electromagnet, Fig. 337. The incandescent lamp is almost completely non-inductive, while the electromagnet has a large amount of self-induction, or a large *inductance*, as it is more commonly called.

If this magnet is connected with a direct current circuit, and if the circuit contains a key, so that the current may be closed and opened, the effect of self-induction in the magnet becomes at once evident in the lamp. When the circuit is closed, the great inductance prevents the current in the magnet from rising at once to its maximum amount. More current, therefore, is forced through the lamp, so that for an instant, that is, for the time during which the current is rising to its maximum value in the magnet, the lamp lights up brightly. As soon as the current in the magnet becomes steady, the counter electromotive force due to self-induction disappears, and the lamp falls in brilliancy and may even be extinguished altogether. When the current through this circuit is broken, the lamp again flashes up brightly, due to the enormous electromotive force of self-induction which now is added to the electromotive force of the circuit.

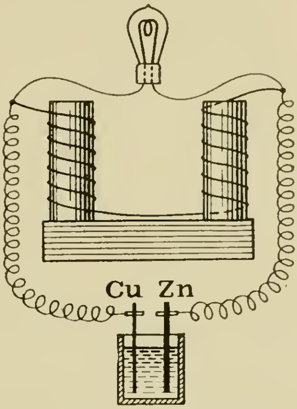

FIG. 337. The counter electromotive force of self-induction.

The lighting up of the lamp in the first instance is indicative of the fact that work is being done in magnetizing the iron, that is, in setting up the strain in the ether all about the magnet, and therefore in producing a magnetic field. When the lamp lights up the second time, the energy which is stored up in the magnetic field, or, in other words, the energy represented by the strained condition of the ether which constitutes the magnetic field, is transformed into the energy manifested by the light and heat of the lamp. It is thus seen that energy is required to produce a magnetic field, and that *a magnetic field itself represents a store of potential energy.*

Since all the lines of force passing through a long solenoid are cut by each turn of the solenoid, the total cutting of lines of force is the product of N, the entire number of lines of force, and n, the number of turns upon the solenoid. N, however, depends upon the strength of the current. Accordingly, *the inductance, L, of a circuit is defined as the total cutting of lines of force when 1 absolute unit current is started or is stopped in* *the circuit.* From this it follows that the inductance depends upon the *shape and configuration* of the solenoid, very largely upon *the presence of masses of iron* enclosed by the circuit, upon *the number of turns of wire, and upon the situation of these turns relative to one another,* thereby determining whether all the conductors or only a portion of them are cut by all the lines of force.

The working unit of induction is the henry, which represents the self-induction of a circuit so constituted that a total cutting of 10^8, or 100,000,000 lines of force is effected when 1 ampere is started or is stopped. Since the electromotive force e of self-induction is the rate at which lines of force are cut,

$$e = \frac{nN}{t},$$

where t is the number of seconds required for the current to rise to its maximum upon closing the circuit, or to fall from its maximum to zero upon opening the circuit. Accordingly, if a current of 1 ampere is broken in a circuit which contains an inductance of 1 henry, and if the duration of this cessation of the current is 1 second, 100,000,000 lines of force will then be cut, and an electromotive force of self-induction equal to 1 volt will be produced.

CHAPTER XLIII.

THE TELEGRAPH AND THE TELEPHONE.

510. The Electromagnetic Telegraph. The electromagnetic telegraph is a device for transmitting intelligible signals by means of making and breaking the circuit of an electromagnet. The Morse instrument consists of a bar of soft iron *A*, Fig. 338, bent into a U-shaped form, and around this a current of electricity may be passed, thus making the soft iron an electromagnet. As soon as this becomes magnetic, it attracts a bar of soft iron *B*, called the armature, placed immediately above it, so that this is drawn down towards the poles of the electromagnet. The armature is pivoted at *C*, and held in position just above the poles of the magnet by the small coiled spring *D*. Accordingly, when the current through the coils of the electromagnet is broken so that this is no longer magnetic, the spring *D* pulls the armature back again into position. The movement of the armature downward coincides, therefore, with the closing of the circuit through the electromagnet, and its movement upward takes place as soon as this current is broken.

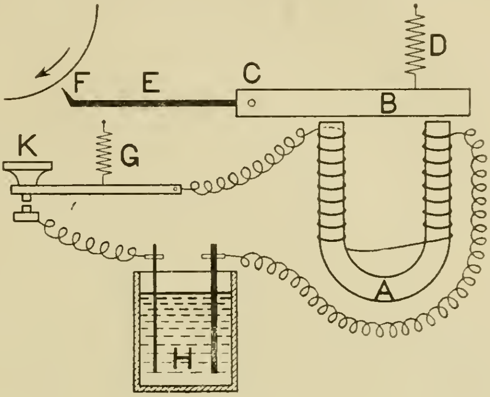

FIG. 338. Diagram of Morse's recording telegraph.

It remains further to devise some means of permanently recording this movement. This is effected by extending the armature by means of an arm *E*, terminating in a pencil point at *F*. Since the armature is pivoted at *C*, the arm *E* and the point *F* are thrown upward whenever the armature is drawn down by the attraction of the electromagnet, and for the same reason these move down whenever the electromagnet ceases to attract the armature and the spring *D* pulls the latter back into position. A cylinder carrying a long strip of paper is now turned uniformly by clock work very near the point *F*, so that a mark is made upon this moving paper whenever the point *F* is thrown upward, that is, whenever the circuit is closed through the electromagnet. Moreover, the length of this mark thus recorded upon the paper is exactly proportional to the length of time during which the current flows, since the point *F* is lowered from the paper as soon as the armature moves upward, that is, as soon as the current ceases to flow in the electromagnet.

Accordingly, when an operator at the distant station closes the circuit at K by pressing with his finger upon a little lever, called the key, the current from the battery *H* flows through this key, around the coils of the electromagnet, making this magnetic. Instantly the armature is drawn down, the point *F* is thrown up against the moving paper, and a mark is recorded. The removal of the operator's finger from the key at *K* enables the coiled spring at *G* to pull the little lever upward, thus breaking the circuit and causing the point *F* to move away from contact with the paper. Therefore, the length of the mark thus recorded upon the paper is absolutely proportional to the length of time for which the circuit is closed at *K*. By closing this circuit for a very short time, simply a dot, or at best but a very short line, is made upon the paper, while closing the circuit for a longer time causes a dash thus to be recorded. Morse, therefore, devised an alphabet in which the letters consist of combinations of dots and dashes. By this means words may be spelled out by making and breaking the circuit properly at *K*, and these signals are thus transmitted to the distant station where they are registered in the same arbitrary alphabet upon the long strip of paper. The paper is then unwrapped from the cylinder and the words again spelled out.

A portion of the modern telegraph instrument, called the sounder, consists of an apparatus constructed exactly like this in every particular with the exception of the recording device, which, in American instruments, is omitted. The movement of the armature up and down makes a series of clicks which the skilled operator learns to recognize, and thus the message is read by sound and at once written down by hand.

511. The Relay. When the telegraph is operated over long distances, the resistance of the wires connecting the two stations becomes so great that the current which can be made to flow through this long, high-resistance circuit is too feeble to operate an instrument such as has been described. Half of the resistance thus introduced by a complete metallic circuit may, indeed, be avoided by using the earth as the return wire, that is, by burying large plates of metal in the earth, one at each station, and letting the current thus flow back to its starting point through the earth which has practically a zero resistance. Even this device, however, does not sufficiently lower the resistance of a long line to enable current enough to flow around the coils of the electromagnet to cause the armature to move with an audible sound.

To overcome this difficulty, an instrument called the relay, or repeater, is used. This consists of an electromagnet A, Fig. 339, wound with several thousand turns of wire and provided with an exceedingly delicate armature, *B*, very carefully pivoted at *C*. The feeble current from the distant station is then able sufficiently to magnetize the iron core of the relay to cause the delicate armature to move, although this movement is wholly inadequate to produce an audible sound.

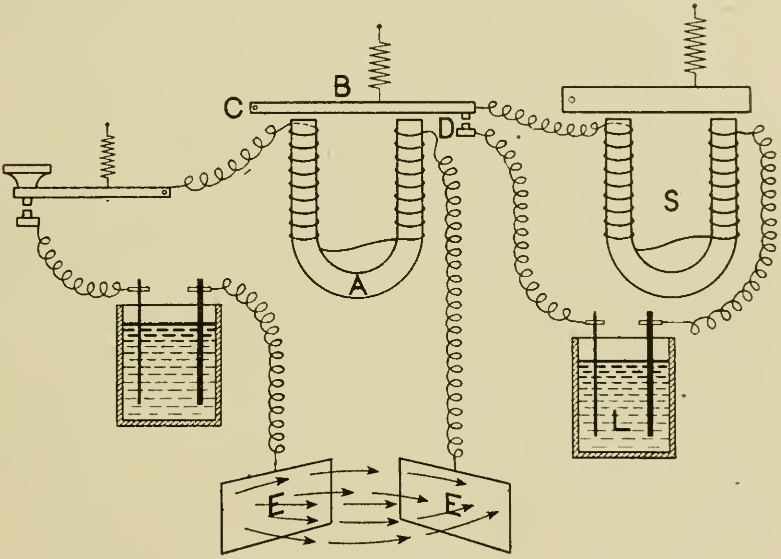

Fig. 339. Diagram of the relay and the local circuit.

Inasmuch, however, as the armature moves, and moves in unison with the making and breaking of the circuit at the distant station, this armature itself may be used as a key which makes and breaks the current in another circuit, called the local circuit, consisting of a battery and a sounder, through which a strong current can flow. As the armature of the relay moves downward, a point upon this armature makes contact with a point *D* immediately below, and this closes the circuit of the local battery *L*. The current from this local battery now flows through the armature of the relay, through the contact at *D*, around the coils of the sounder *S* which is wound with heavy wire, and so back to the battery *L*. A strong current flows through this low-resistance local circuit, making the electromagnet of the sounder strongly magnetic, and bringing the armature down with a sharp click, which is readily heard. In this way the relay really repeats the message automatically by opening and closing the local circuit simultaneously with the opening and closing of the line circuit at the distant station.

512. The Cable. In attempting to send a message through the cable across the ocean, it is found that the current thus transmitted is too feeble to affect even the most delicate relay. This is due not only to the resistance of the cable, but in a far greater degree to the delay which the current experiences in reaching its full intensity, because of the *capacity* of the cable. The copper wire forming the true conductor of the cable is separated from its iron sheathing and the sea water by a layer of insulation, thus making it, indeed, a very long condenser of very large capacity. So great is this capacity that an interval of several seconds must elapse after the current is sent into the cable at one end before it rises to its maximum intensity at the other, and even then it does so only gradually.

On this account the receiving instruments are made exceedingly delicate, so that they will respond when the current begins to rise, and thus shorten the time of transmission. Moreover, the cable is not charged to its full capacity for each signal, but is given a charge for an instant only, and is then connected with the earth, or "grounded", so that an electrical pulse, or wave, is all that actually advances. Such a pulse sent from the positive pole of the battery represents the dots of the telegraphic code, and a corresponding pulse from the negative pole indicates the dashes. To avoid disturbances from earth currents, which are nearly always present and which would mask the signals transmitted, each end of the cable is entirely insulated, and is connected to one side of a large condenser, while the charging battery and the receiving instrument are connected between the other side of the condenser and the earth.

In place of the relay for receiving the message, there is substituted an exceedingly delicate galvanometer, and the message which is being received is read from the deflections of the needle. Formerly, this needle carried a mirror from which light could be reflected, and the message was then read by the movement of this spot of light to the right or left along a scale in a darkened room. Lord Kelvin, however, devised a little siphon pen so delicate that it is mounted upon this very light needle, and, of course, turns with the needle. A strip of paper drawn beneath this pen has, therefore, a wavy line recorded upon it, the deviations of this line, first in one direction and then the other, corresponding to the dots and dashes of the Morse alphabet which is used.

The conductor of the cable consists of several strands of copper wire twisted together, thereby insuring both flexibility and working service, even if a number of the strands should break. This copper conductor is most carefully surrounded by many layers of hemp saturated with gutta percha, which protect it absolutely from contact with the sea water. The hemp in turn, is surrounded by a sheathing of steel wires to prevent the cable from being cut by sharp rocks on the bottom of the ocean, and also to give it sufficient strength to support its own weight while being laid. So accurately are electrical measurements now conducted, that when a "fault" occurs, either when the cable parts altogether, or when the insulation has been injured, thus admitting the conducting sea water to the copper wire, the position of this break may be readily located. Ships therefore sail out to the place where the fault occurs, the cable is picked up by grappling hooks and the injury repaired, after which it is once more lowered to its position upon the bottom of the ocean.

513. The Telephone. The electromagnetic telephone, as devised by Alexander Graham Bell, consists, in its essential parts, of a permanent steel magnet, Fig. 340, with a coil of fine wire wound around one end, and provided with a thin, soft iron disc, very near indeed to this end of the magnet. With this apparatus no battery is used. The lines of force of the magnet pass into the soft iron disc, since this has a high permeability, and flow through this disc to its edge, thence through the air to the other end of the magnet. This entire flow of lines of force takes place through the coil of wire which is wound around the end of the magnet. When the disc is made to approach nearer to the magnet, it intercepts a greater number of lines of force, and thus makes the induction greater. The field of force through the coil of wire is thus increased, and this develops an induced current in this wire. When the disc is moved away from the end of the magnet, not so many lines of force pass through it, and consequently the induction is lessened. This weakening of the field of force within the coil also induces a current, but in the direction opposite to that of the former current.

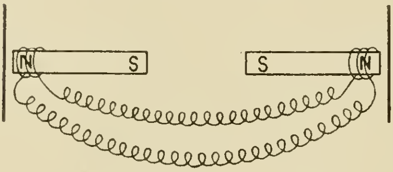

FIG. 340. Diagram of the telephone.

When speech is uttered before this device, the vibrations of the air throw the disc into rapid oscillations, causing it to move back and forth very near the end of the magnet. The approach of the disc develops a current flowing in one direction through the coil of wire, while the recession of the disc induces a current flowing in the opposite direction.

If the two ends of this coil are now connected with the two ends of a similar coil in a precisely similar instrument, the current which is generated in the first coil flows also through the second. As the current flows in one direction around the coil of the second instrument, it strengthens the magnet a little, so that this attracts the diphragm a little nearer; and as the current flows in the opposite direction around this coil, it weakens this magnet to some extent, so that the diaphragm, not being so strongly attracted, moves a little farther away from the end of the magnet. In this way the diaphragm of the second instrument is thrown into vibrations precisely similar to those which the first diaphragm executes, and these are similar in every way to the vibrations of the air produced by the voice. The second diaphragm, therefore sets the air in its neighborhood into vibrations which coincide in all respects with the vibrations occurring before the first diaphragm; and therefore, by holding the second diaphragm to the ear, sounds are heard similar to those which were uttered before the first. It is, therefore, perfectly possible to use this instrument both for transmitting as well as for receiving sounds.

514. The Microphone Transmitter. Another mechanism, however, is now used as the transmitter, or as the generator for the alternating and fluctuating currents which are sent along the line. This device depends upon the principle that the electrical resistance of a bad contact varies greatly with the pressure to which it is subjected, so that a current which is sent through such a poor contact increases as the pressure increases, and diminishes as the pressure diminishes.

Application is made of this principle in the construction of the modern transmitter of the telephone. To the diaphragm of this transmitter is attached one-half of a little box *A*, filled with granular carbon, Fig. 341, so that when the diaphragm moves back and forth in response to the vibrations of the voice, the pressure of the little particles of carbon upon one another varies continually. If the current from a battery *B* is made to flow through the carbon in the box, the varying pressure due to the vibrations of the diaphragm causes the resistance of the carbon to vary. This variation of the resistance causes the current which flows through this carbon continually to change, the variations in the intensity of the current agreeing with the vibrations of the diaphragm and depending upon them. This current of varying intensity is then sent around the primary of a little induction coil, while the current which is thus induced in the secondary is the current that is sent along the line to the distant instrument. This fluctuating current passes around the coil which surrounds the magnet of the hand telephone, or the receiver, and thus reproduces the sound in the method already described.

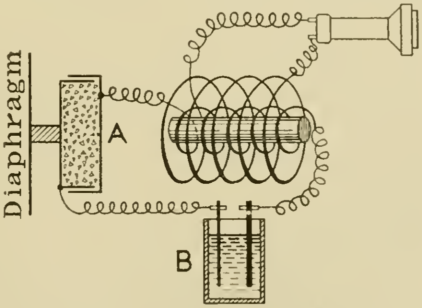

FIG. 341. The microphone transmitter and its connections.

This principle, whereby a feeble vibration is made to vary the electrical resistance of a bad contact and so produce an audible sound in the telephone, is known as that of the microphone. If a carbon rod *A*, Fig. 342, is laid loosely upon a carbon plate *B*, and if a current from a battery flows through the poor contacts made by these pieces of carbon in touching each other, and then flows through the telephone, the least vibration occuring in this carbon system will cause a sound to be heard in the telephone. Indeed, if ordinary conversation is carried on before this apparatus, it will be heard in the distant telephone, or if a watch is laid upon the table where this instrument stands, its ticking may be distinctly heard.

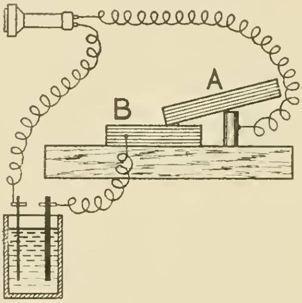

FIG. 342. The principle of the microphone.

515. Electric Waves and Wireless Telegraphy. The year 1887 marks an epoch in electrical science, for then was ushered in a new era of knowledge and investigation in electricity. In that year Heinrich Hertz, of Carlsruhe, Germany, made the important discovery that the disturbance in the ether produced by an electric spark is propagated through space with the velocity of light and moves with wave motion. These ether strains or electromagnetic waves are again converted into electrical currents whenever they meet conducting circuits, so that throughout the intervening space the electrical energy is propagated in waves, as is the case with light. These waves are called *electromagnetic*, because they have an electric and also a magnetic component.

In our study of electrical discharge we have seen (p. 402) that the spark is not a simple passage of electricity from one conductor to the other, but consists of a series of a hundred or more oscillations first one way and then the other. A single spark, therefore, consists of many oscillatory discharges, each oscillation occupying seldom more than the millionth part of a second, and often less than the thousandth part even of this small interval. With each oscillation, the ether in the immediate vicinity of this spark is strained (p. 323), so that the strain in the ether, as well as the spark, is oscillatory in character. Like other vibrations in the ether, these electromagnetic waves are propagated with the velocity of light, and the distance which the disturbance travels during the interval of one complete oscillation of the electric spark is known as the wave length. The period of these oscillations, however, depends upon the capacity and the self-induction of the circuit producing the spark, and therefore the wave length as well as the period of the ether waves depends also upon the capacity and self-induction of the sending circuit.

Hertz detected the presence of these electromagnetic waves throughout a space ten or fifteen meters from the source by placing a loop of wire a little more than half a meter in diameter in different positions relative to the spark, and observing the presence or absence of tiny induction sparks between the two ends of this loop when these were separated by an exceedingly small fraction of a millimeter. In this way he detected the existence of nodes and antinodes, measured wave lengths, and demonstrated that these electric waves may be reflected, refracted, and polarized, exactly as is the case with the ether waves forming light. The researches of Hertz have proved beyond a doubt the identity in nature of light and electric waves, and have furnished a striking experimental verification of the theory propounded by Maxwell twenty years earlier, that light is an electromagnetic phenomenon.

516. The Coherer. Subsequent to the discoveries of Hertz, Branly found that the electrical resistance offered by a tube of metallic filings decreases enormously when exposed to electric waves. If a small tube containing metallic filings, preferably of nickel and silver, is included in series with a battery and a relay, the resistance offered by the filings is so great that the relay fails altogether to act, and, indeed, practically no current flows. As soon, however, as electric waves sweep across this tube, the resistance becomes very small, so that a current then flows through the relay, and this, in turn, actuates a sounder. A slight tap with the finger restores the filings to their original high-resistance condition, the current then ceases to flow through the relay, and the apparatus is ready to detect another electric wave.

The action of the electric waves upon the tube of filings is not well understood. It was formerly thought that tiny induction sparks passed between the filings, causing them to weld together, or to *cohere*, thus forming a circuit of low-resistance, and the tube of filings was therefore called the *coherer*. This name is still retained, although it is now thought that the electric waves develop a difference of potential between the filings which causes them to exert so powerful an electrostatic attraction for one another that the layer of air between them is excluded, and the filings come into intimate contact. An automatic tapping device is arranged whereby the coherer may always be de-cohered as soon as the wave passes, and thus be in a condition to detect succeeding waves. A convenient means of accomplishing this is to mount the coherer upon the armature of the sounder which the relay actuates, so that the filings are jarred, decohered, and therefore ready for the next electric wave as soon as the signal due to the presence of the first one has been given.

517. The Aerial. It was soon discovered that, by connecting a long vertical wire with one of the two conductors between which the oscillating spark passed, more energy may be thus imparted to the ether, and also that by connecting the coherer, or other receiving apparatus, with a similar vertical wire or aerial, a greater amount of energy may be received. Accordingly, both at the sending and at the receiving stations of the wireless telegraph, tall towers or masts are erected from which these aerials, or antennae, are suspended. In these aerials are developed induction currents which vibrate throughout the length of this wire in unison with the electromagnetic waves which they intercept. Originally the coherer was inserted between the aerial and the earth, so that these induction currents could pass through it. Later, it was found that a more efficient method was to allow these oscillating currents in the aerial to pass through the primary of a small induction coil which contains no iron, and to connect the coherer in the secondary circuit of this coil.

518. Wave Detectors. Many other means for receiving the electric waves have been devised, all of which may be grouped together under the general term of "detectors." One in particular, known as the magnetic detector, deserves mention, inasmuch as it has been used with peculiar success by Marconi in his transatlantic telegraphy. In this device, the current from the aerial passes through the primary of a small induction coil, the core of which is a ribbon of soft iron magnetized to saturation. This iron, by its great retentivity, retains a large amount of residual magnetism, though this is exceedingly unstable, since the coercive force of soft iron is very small. The secondary of this induction coil is connected with a telephone, which will thus indicate any change in the number of lines of force passing through the primary, that is, it will indicate any change in the magnetic condition of the soft iron ribbon. The oscillating current from the aerial, as it passes through the primary of this coil, acts as a coercive force, which destroys the residual magnetism of the iron ribbon. This decrease in the number of lines of force passing through the iron core induces a current in the secondary coil, and a sound is produced in the telephone. A fresh part of the magnetized iron ribbon is then moved by clock work into the primary of this coil, so that the following waves may be detected in a similar manner. The message is thus read by a succession of clicking sounds in the telephone.

Another detector which has been found effective and useful depends upon the property which many crystals possess of rectifying an oscillatory current. If an alternating current is made to pass through the contact formed, for instance, by a crystal of carborundum and a metallic plate, it is found that the part of the current which flows in one direction will pass freely through such a contact, but that the portion which flows in the opposite direction cannot pass. In other words, the contact formed by a crystal of carborundum and a metallic plate will allow a current to pass which is flowing in one direction, and this combination of carborundum crystal and metallic plate is accordingly called a rectifier.

If such a rectifier is connected between the aerial and the earth, only those portions of the oscillating current in the aerial which flow in one direction will pass through it, while those oscillations which take place in the opposite direction are excluded. Accordingly, if a telephone is connected in series with the rectifier between the aerial and the earth, those portions of the current which pass through the rectifier will likewise pass through the telephone and produce a sound, and thus the message may be read. Another and better method is to connect the telephone as a shunt to the rectifier, in which case those portions of the current which are excluded by the rectifier will flow through the telephone and thus produce a sound. In practice, the telephone is now used almost altogether for receiving wireless messages.

As early as December 1901, Marconi was able, by using the magnetic detector, to recognize the three dots of the letter "S" which were transmitted by ether waves across the Atlantic Ocean. During the succeeding year he made many improvements in his apparatus; and early in the year 1903, a message was sent without wires across the Ocean from President Roosevelt in the United States to King Edward in England. Later, a regular transatlantic service between these two countries was established, and in 1914 wireless communication between Berlin and New York was made possible.

CHAPTER XLIV.

CONDUCTION OF ELECTRICITY THROUGH GASES.

519. The Ionization of a Gas. When two conductors in the air are brought to potentials which differ by several thousand volts per centimeter, an electrical stress of such magnitude may be developed that the air will be broken down and a spark will be produced. This spark is a true passage of electricity, and a momentary current flows from one conductor to the other. Under ordinary conditions, however, and for moderate differences of potential, dry air at atmospheric pressure is found to be practically a non-conductor of electricity, so that a charged electroscope, for instance, may remain often for hours without having its charge perceptibly diminished.

Nevertheless, by a number of processes, the air may become a fairly good conductor, and when in this condition it enables electricity to flow continuously between two conductors, or electrodes, even though these be maintained at potentials which are relatively low. The methods by which the air may be made conducting are numerous and varied, such as the splashing of near-by water, the gases from flames, the presence of incandescent metals, the action of ultra-violet light upon polished metallic surfaces, the previous passage through this air of an electric spark, the action of the X rays, the presence of radium or any radio-active body, and to some extent a spontaneous action from the air itself. Under any of these conditions except the last, an electroscope, whether charged positively or negatively, will be rapidly discharged. Air when in the conducting condition is said to be *ionized*.

The conductivity thus imparted to air or any other gas may be removed by drawing the air or gas through a tube closely packed with cotton or with glass wool, thus showing that the cause of the conductivity is due to particles which may be removed by the mechanical process of filtering. This conductivity may likewise be removed by the presence of an electric field, which shows that these particles are electrified, and are thus driven by the field against the sides of the tube in which the gas is contained.

Since the gas as a whole is neutral, that is, it is charged neither positively nor negatively, it is evident that there must be present both positively and negatively charged particles in equal numbers. Therefore, it is seen that the conductivity of a gas is due to the movement of both positively and negatively charged particles, which are known as the *ions*, though these ions are not identical with those to which the conductivity of electrolytes is due. The presence of these ions in the gas suggests a molecular dissociation, and, indeed, this is found to be the case. Any one of the processes above mentioned is able, therefore, to separate a single, free, negatively charged electron from the gaseous molecule, and thus leave the latter positively charged, the two charged parts being called the *ions*, and this process being known as the *ionization* of the gas.

520. The Geissler Vacuum and the Crookes's Vacuum. When the pressure of the gas is decreased, its character, as regards its conductivity, changes completely. This may be conveniently studied by sealing two wires, or electrodes, into the ends of a glass tube from which the air may be gradually exhausted. If a difference of potential far less than that required to produce a spark is maintained between these electrodes, but little effect will be observed, until a pressure equivalent to 7 or 8 centimeters of mercury is reached, when a beautiful, violet-colored thread, or band of light, may be seen extending from one electrode to the other, and a continuous current of electricity now flows through the tube. As the exhaustion proceeds, the conductivity apparently undergoes a modification, for the appearance of the tube changes altogether, the violet band becomes broader and paler, while a dark line develops in the immediate neighborhood of the negative electrode, or the cathode.

A very characteristic stage is reached when the pressure of the gas is reduced to one or two millimeters of mercury, that is, to the condition which is known as the Geissler vacuum. It will then be observed that a faint, luminous, velvety coating, *A*, Fig. 343, surrounds the cathode, and that the line near that electrode has widened, forming a dark region, *B*, known as the Crookes's dark space. Beyond this again is observed a faint luminous patch, *C*, known as the negative or cathode glow, after which comes another dark region, *D*, called the Faraday dark space, and adjacent to this is a highly luminous band of light known as the positive column, *E*, often forming into sections or striae of unequal intensity, and extending clear to the positive electrode. It is worthy of note that the length of the Crookes's dark space is independent of the length of the tube, or of the position of the positive electrode, or anode. In a tube fifty feet long, the luminous positive column occupies the entire distance with the exception of an inch or two immediately beyond the cathode.

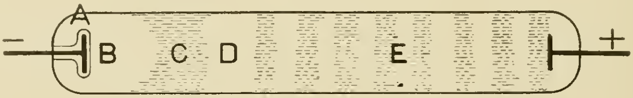

FIG. 343. The Geissler discharge.

As the exhaustion of the tube is carried still farther, the resistance becomes greater, the Crookes's dark space increases in length, the positive column becomes less luminous and shorter, and seems to be pushed back towards the anode as the Crookes's dark space becomes longer. When the pressure in the tube has been reduced to about .001 millimeter of mercury, the positive column disappears, the Crookes's dark space increases in length until, for tubes of moderate size, it occupies the entire tube, and the glass at the end remote from the cathode emits a bright green fluorescent light. When this stage of exhaustion has been reached, a degree of rarefaction known as the Crookes's vacuum has been produced; and the tube in which a discharge of this nature occurs is known as a Crookes's tube.

521. The Cathode Rays. It is from the discharge which takes place through a tube at this low pressure, and from many of the phenomena which may be observed in connection with this discharge, that conclusions have been reached as to the nature of electricity, and deductions drawn regarding the constitution of matter. When we recall that a gas is able to conduct electricity only when it contains a number of electrified particles mingled with its molecules, it is evident that in the Crookes's tube we are dealing with a stream of such particles. Sir William Crookes, who first studied a gas under these conditions, showed that an obstacle placed between the electrodes would cast a shadow on the end of the tube away from the cathode, thereby proving that this conduction was effected by particles which were shot out from the negative electrode, and which are now known as the cathode rays. This cathode stream was observed to be deflected when an electrified body was brought near, and was deflected in such a direction as to show that *these cathode rays consist of negatively charged particles.* Moreover, when a magnetic field was approached, this steam of cathode rays was again deflected, but in a different direction, moving, indeed, at right angles to the stream, and also at right angles to the magnetic lines of force, that is, moving in a direction in which a negative current of electricity would move if it were flowing in this magnetic field. It was seen, therefore, that *this stream of negatively electrified particles, moving through* *the tube from the cathode to the anode, constitutes the current of electricity flowing through the tube.* Even at this early date, 1883, Crookes, by an almost prophetic insight, predicted that the particles of which the cathode rays consist would be found to be smaller even than the atoms.

522. The Velocity, Charge and Mass of the Cathode Particles, or the Electrons. Much of the knowledge which the world now possesses regarding the nature both of electricity and of matter, and of the relation between them, is due to the untiring work of Sir J. J. Thomson, and to the researches of the persons whom his teachings have inspired.

By using an exhausted tube of a form similar to the one shown in Fig. 344, where *C* is the cathode and *A* the anode,

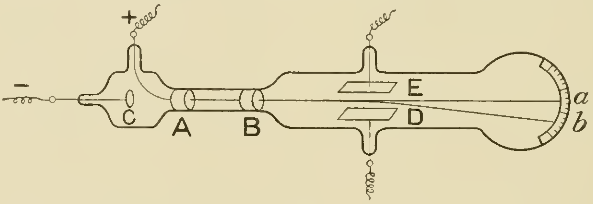

FIG. 344. Tube for measuring the deflection of cathode rays by magnetic and electric fields.

Thomson was able to measure not only the velocity, *v*, of the cathode particles, but also the ratio of the charge, *e*, which each particle carries, to the mass, *m*, of that particle. The anode, *A*, is pierced by a hole perhaps a millimeter in diameter, and *D* is a metallic plug connected with the earth, and having a similar hole at its center. Cathode rays, after passing through the holes in *A* and *B*, continue in this direction through the tube, and at the farther end fall upon a fluorescent screen, which becomes luminous under the impact of these rays.

By applying a known magnetic field of intensity *H*, this beam of cathode particles will be deflected through a curved path, the radius, *r*, of which may be computed from the distance *ab*, which is the amount of deviation produced by the field. But when a magnetic field, *H*, acts at right angles to a current, *I*, it exerts upon the current a force equal to *H I*. In this case, the current which flows is produced by particles, each charged with *e* electromagnetic units of electricity, and moving with a velocity of *v* centimeters per second. Each moving particle, therefore, is equivalent to a current *ev,* and the force, due to the magnetic field, *H,* tending to deviate this particle from its rectilinear motion, is $H e v$. Inasmuch as the particle moves in a curved path, the deflecting force, acting upon it in the direction of the radius of curvature, must be equal to the centrifugal force, $\frac{m v^2}{r}$, of the particle, acting in the opposite direction, and therefore

$$H e v = \frac{m v^2}{r},$$

from which

$$H r = \frac{m v}{e}. \qquad (1)$$

Since H and r may both be measured, the value of $\frac{m v}{e}$ may be determined.

At D and E in the narrow part of the tube are inserted two metallic plates which may be connected with a battery, and thus a difference of potential E produced between them. The negatively charged particles, of which the cathode rays are composed, are therefore repelled by one plate and attracted by the other, so that this beam is deflected likewise by this electrostatic field. The force acting on any particle tending to deflect it is proportional to the intensity of the field E and to the charge e which the particle carries, hence this deflecting force is equal to Ee. By so adjusting E that its value and direction will be such as to produce a deflection of the cathode beam exactly equal in amount and opposite in direction to that produced by the magnetic field, each force will hold the other in equilibrium, and the luminous spot upon the phosphorescent screen will remain undeviated. Accordingly,

$$H e v = E e,$$

and

$$v = \frac{E}{H}.$$

The velocity of the cathode particle is thus found to be simply the ratio of the intensities of the electrostatic and the magnetic fields, both of which may be measured, and hence v is known. The value of v, determined in this way, is found to be from $\frac{1}{5}$ to $\frac{1}{2}$ the velocity of light, depending upon the difference of potential between the electrodes, that is, the cathode particles move with a velocity from 40,000 to 90,000 miles per second. By substituting the value of v thus obtained in equation (1), the value of e/m may be computed, since both H and r are known. The value of e/m thus determined is 1.7×10^7, where e is the charge upon each cathode particle measured in electromagnetic units, and m is the mass of this particle.

By the extremely beautiful and ingenious device of holding in equilibrium between two charged plates a drop of oil to which an electron has attached itself, Millikan of Chicago has measured the electrical charge e upon a single electron, and found it to be 4.77×10^{-10} electrostatic units. *This, however, is identical with the charge which is carried by the hydrogen ion in electrolysis.* Accordingly, by comparing the value of e/m as thus determined for the cathode particle, namely, 1.7×10^7, with that as found for the hydrogen ion in electrolysis, which is almost exactly 1×10^4, we find that e/m for the cathode particle is 1700 times greater than the value of e/m as determined for the hydrogen atom; and, therefore, since e is the same in both cases, it follows that m, *or the mass of the cathode particle, is only* $\frac{1}{1700}$ *of that of the H atom.*

523. The Electron is the Ultimate Unit or Atom of Electricity. In this way Sir. J. J. Thomson proved that the atom is not the ultimate particle of matter, but that it is extremely complex, and contains in its constitution a number of these exceedingly small particles, no larger than the $\frac{1}{1700}$ part of size of the hydrogen atom, which are called *corpuscles* or *electrons.*

Probably no one discovery of modern science is of greater import than this, or is attended with consequences which are farther reaching.

Moreover the nature and characteristics of the electrons which are present in a beam of cathode rays are found to be absolutely independent of the material of the electrodes, or of the kind of gas which the tube contains. It is therefore apparent that these electrons are a common constitutent of matter of all kinds, and that matter is composed, in part at least, of these negatively charged particles, together with a certain amount of positive electrification, about which but little is now understood.

Not only is the electron the smallest particle of matter known to exist, but its electrical charge, which is constant, is the smallest quantity of electricity that has ever been detected. Indeed this quantity of electricity is so small that none has ever been found smaller. Again, larger quantities of electricity are found always to be multiples by whole numbers of the quantity upon an electron, from which it is apparent that this is the ultimate charge or unit of electricity, that a negative static charge is a multitude of these electrons in undirected motion upon the surface of the charged body, and that an electric current is simply an assemblage of these electrons moving with a directed motion.

524. The Electrical Theory of Matter. A consideration of the greatest interest connected with the electron is the possibility that this particle, although obtained from matter of all kinds, may itself not be composed of *matter*, as this word is ordinarily used, but that *its mass may be only apparent,* and due to the fact that it is an electrified particle moving with a high velocity. It is well known that a charged body, when set in motion, spreads a magnetic field, and therefore also electromagnetic energy, throughout the surrounding region. Accordingly, more force is required to accelerate a charged body than to set in motion a body which is not charged, so that a charged body has a greater apparent mass than one uncharged.

It is thought that the same reasoning must be true with the electron. Indeed, by using electrons produced by radium, Kaufmann has detected indications that the mass of the electron is not constant, but increases with its velocity. Moreover, Thomson has computed the apparent, or the electromagnetic, mass of the electron from its charge and its velocity, and has found it to agree with its mass as determined by experiment. If these conclusions prove to be correct, we may regard the electron simply as the ultimate atom or particle of electricity, and as having mass which is only apparent, that is, which is wholly electromagnetic and not material. Moreover, since electrons may be obtained from matter of all kinds, the question may fairly be asked if all matter, or at least a portion of it, may not be entirely electrical in its constitution.

525. Conduction of Electricity Through Gases. We are now in a position to discuss more intelligently the phenomena which occur when the electrical discharge takes place through the rarefied gas of the ordinary vacuum tube, (p. 634). Observation has shown that, due to the spontaneous ionization before mentioned, the air or any gas ordinarily contains a few ions, that is, a few free electrons and a corresponding number of air molecules each of which has lost an electron, and which is therefore positively charged. When a sufficiently great difference of potential is produced in the air at atmospheric pressure, these free electrons are repelled by the negative electrode, and collide with a near-by molecule with sufficient energy to dislodge an electron from it. This electron is then likewise repelled by the cathode, and collides with another molecule from which it displaces an electron, and so on, and a spark passes.

When the pressure is reduced, as it is in the exhausted tube, the molecules are farther apart, the electron can then move through a longer distance before colliding with a molecule, and does so with a much smaller difference of potential between the electrodes. When the pressure is no lower than that of six or eight centimeters of mercury, and the violet-colored thread is observed to extend between the electrodes, this process of collision and ionization goes on throughout the length of the tube, and the current thus flows from one molecule to the next by the movement of an electron that was dislodged from one of the previous molecules. Whenever a molecule is thus bombarded by an electron, or whenever a molecule which is lacking an electron picks one up, or recombines, a shock of such intensity is thereby imparted to the electrons remaining in the molecule that they are made to vibrate; and as their vibration means the periodic movement back and forth of electrified particles, electromagnetic waves will go out from this molecule and produce the sensation of light. The violet color of this thread of light is characteristic of nitrogen, and is due to the vibration of the electrons in the nitrogen molecules which are thus bombarded by electrons.

But the cathode not only repels negatively charged electrons, but attracts the positively charged ions that have lost an electron. These larger particles, thus being attracted by the cathode, move toward it with so high a velocity and strike it so severe a blow that they dislodge an electron from the metal of which it is made, and this electron is then repelled by the cathode in a straight line perpendicular to its surface. The velvety light immediately about the cathode is thought to be due to the positive ions which are here bombarding this electrode, and the Crookes's dark space is the region through which are moving the electrons which had been repelled in straight lines from its surface. The somewhat complicated distributions of light throughout the various parts of the tube are thought to be the places where either ionization or recombination is taking place, and where a shock is thus imparted to the electrons in the molecules. As the exhaustion increases the number of molecules diminishes, and so the collisions occur less frequently. The Crookes's dark space then grows longer and the positive column shorter, until, at a pressure of .001 millimeter or less, the positive column disappears, and the Crookes's dark space extends clear to the annode and to the glass behind it.

526. The Energy of the Cathode Rays. Many beautiful and most interesting phenomena are connected with the passage of this column of cathode rays throughout the length of the tube, and the effect produced by the impact of these rays upon various objects is likewise full of interest. As has been pointed out, the glass then glows with a bright fluorescent light, the color of which depends upon the materials of which the glass is made. An object placed in the tube casts a perfectly sharp shadow upon the remote end, showing that the cathode rays are shot out perpendicularly from the cathode, and move in straight lines. Indeed, the path of these rays through the tube is entirely independent of the position of the anode, for the column always proceeds in a straight line from the surface of the cathode.

The impact of the cathode particles produces the most beautiful and brilliant fluorescence in various crystals which may be placed in the path of these rays. If the cathode is made concave, so that the rays which leave its surface will be brought to a meeting point, or focus, and if a plate of platinum is placed at this focus, the energy of these rays, due to their enormous velocity, will raise the platinum to a bright red heat. Indeed, the kinetic energy per milligram of the electrons constituting the cathode rays has been calculated as 10^{17} ergs, or about three and a half million foot-tons.

527. The Rontgen Rays. But perhaps the most interesting phenomenon connected with this discharge is the fact that any surface which is thus bombarded by these cathode particles becomes itself the seat of a new form of radiation, discovered by Professor Röntgen, of Munich, and now known as the Röntgen rays, or the X rays. Formerly the source of these rays was therefore the inside of the tube itself, as the stream of cathode particles impinged upon it. In the modern tube, the cathode rays are brought to a focus by a concave electrode, as described above, and the Röntgen rays thus emanate from the platinum target which is placed in this focus, Fig. 345.

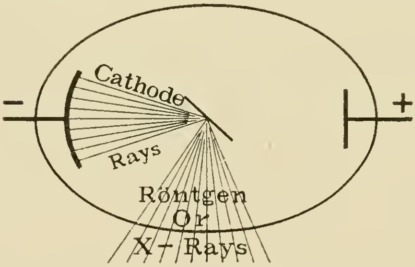

FIG. 345. The modern Crookes's tube, showing the production of the X rays.

This new radiation, which not only penetrates may opaque objects, but which also acts photographically, passes through the vacuum tube, or the Crookes's tube, as it is now called, and thus may be detected at wide distances from its source. While these Röntgen, or X, rays are perfectly invisible, they have the power of causing a fluorescence in certain chemical substances, notably in platinum barium cyanide, and a screen coated with this salt, when exposed to these rays, becomes luminous with a faint, greenish-yellow light.

Dewar found that the relative opacity of different substances to these rays is approximately proportional to their atomic weights, so that a body of low atomic weight, such as carbon, is extremely transparent to this radiation, while bodies of higher atomic weight, such as nearly all the metals, are far more opaque. It is possible, therefore, to photograph metallic objects enclosed in a wooden box, or to have these objects cast a deeper shadow upon a fluorescent screen, for these rays pass almost unimpeded through the carbon of which the wood is composed, while they are partially, if not entirely, intercepted by the denser metal. In a similar way, because the bony structure of the body is denser and composed more largely of metallic salts than the fleshy portions, the bones cast a deeper shadow upon a photographic plate or a fluorescent screen, and in this way they may be examined with almost as great facility as if the flesh were not present.

Of the many theories that have been framed to explain the nature of the X rays, perhaps the most plausible is the one due to Stokes, which assumes that the impact of the electrons upon any surface whatever gives rise to ether waves of far higher frequency, and therefore of shorter wave length, than light, but not periodic, since they have been produced by shocks, or impulses, rather than by regular vibrations. Indeed, this stoppage of the electrons as they impinge upon the platinum plate or the walls of the tube, produces a sudden change in the magnetic and electric fields which accompany the swiftly moving, negatively charged particles, and therefore ether waves must go out from these points of arrest as centers. In support of this theory, Thomson has suggested that the fluorescence of the walls of the tube and of many substances bombarded by the cathode particles may likewise be due to the sudden change thus produced in the electric and magnetic fields of the electrons.

CHAPTER XLV.

RADIO-ACTIVITY.

528. The Discovery and Isolation of Radium. Soon after Röntgen's discovery, the idea occured to Becquerel that there might be natural substances which would of themselves give out a radiation similar to that produced by the cathode rays. After trying many kinds of matter, he found that metallic uranium, as well as the salts of this element, after exposure to sunlight would produce effects not unlike those caused by the X rays. Later it was found that the sunlight played no part in this process, but that this result was due to an unknown kind of radiation which emanated naturally from uranium and its compounds.

Further investigation showed that certain ores of uranium, notably pitch-blende, produced this effect even more energetically than pure uranium; and from this fact the conclusion was apparent that these ores contained traces of one or more elements of still greater activity than uranium itself. A long and remarkable chemical analysis of a ton of pitch-blende was then undertaken by M. and Mme. Curie, the result of which was the isolation of a few milligrams of two exceedingly active substances, one of which Mme. Curie named polonium, after Poland, the land of her birth, and the other, two million times more active than uranium, which she called radium. This substance was indeed fortunately named, for it was found to be continually emitting a radiation which has the power of ionizing the air, exciting fluorescence in many substances, and acting photographically upon the sensitive plate.

529. Character of the Radiations from Radium. Prominent and foremost among the investigators who began at once to study this new phenomenon, was Sir Ernest Rutherford, to whom the world is mainly indebted for its knowledge of this subject. By placing a small quantity of radium in a leaden block through which these radiations could not pass, Rutherford was able to obtain a beam of these rays which were emitted from an opening at the top. When this block was brought between the poles of a powerful magnet, it was observed that these rays were deflected, thereby showing that they were produced by charged particles of matter which the radium was constantly throwing out or expelling from itself. It was evident, therefore, that the activity of radium is due to a process of the disintegration of this material, and that this element is unceasingly expelling from itself charged particles, which either formed a part of its own mass, or which entered into the composition of its molecule.

Moreover, from the direction and the magnitude of these deflections, Rutherford concluded that there were two very different kinds of matter thus expelled, one, a positively charged mass which he called the α particles, having an atomic weight of 4, and moving with a velocity about 1/15 that of light, or 12,000 miles per second, and the other, a negatively charged body which he called the β particle, having a mass $\frac{1}{1700}$ that of the hydrogen atom, and moving with a velocity very nearly equal to that of light, or 180,000 miles per second. Besides these two varities of charged particles, he found that there was a third type of radiation which was entirely undeviated and uninfluenced by the magnetic field, and to this he gave the name of the γ rays.

The keenest interest was then aroused as to the nature of these three radiations emitted by radium. From the fact that the mass of the α particle is 4, which is the atomic weight of the element helium, it was early suspected that the α particle was simply a helium atom, and this has subsequently been proved to be the case. When emitted from radium, it is, however, charged positively with two unit charges, or twice the charge carried by the hydrogen ion, and from this it is apparent that the α particle is a helium atom that has lost two electrons. The β particle, on the other hand, is found to be charged negatively with a single unit charge, that is, with the charge of a single electron, and it likewise has the same mass as well as the same velocity as the particles which we have found to constitute the cathode rays. Accordingly, the characteristics of the β particles are in every way the same as those of the cathode particles in the Crooke's tube, and these particles are therefore believed to be identical with the cathode particles; or, in other words, the β particles are simply rapidly moving electrons.

This conclusion as to the nature of the β particles renders simple the interpretation of the γ rays, which invariably accompany the emission of rapidly moving β particles, and which seem to be neither corpuscular nor electrified, but to have the character of a true radiation. When it is remembered that any substance bombarded by cathode rays, or electrons, gives rise to X rays, and when we recall that the β particles have thus been shown to be simply electrons, we see that the β particles. in striking against the walls of the vessel and the substance of the radium itself, must therefore give off X rays; and accordingly the γ rays are believed to be identical in nature with the X rays.

530. Radio-Activity Due to Atomic Disintegration. From what has been said as to the nature of radio-activity, it is evident that the processes which are taking place in radium, or in any radio-active body, are such as to cause a disintegration of the substance. It has been seen that, as α and β particles are thus projected, the radium itself must continually be losing in weight. Moreover, when we consider that the atomic weight of radium is 226.5, and that the mass of each projected α particle is 4, while that of the β particle is only $\frac{1}{1700}$, it is evident that it is not molecules of radium that are thus being thrown out from the quantity present, but that *radio-activity is an atomic process and that it is the atom of radium itself that is thus disintegrating.* The conclusion is also apparent that through the continual loss of particles from its own material, a quantity of radium is constantly becoming less, and that in time it must disappear altogether. In computing this time, the loss of mass due to the expulsion of the α particles alone is considered, since the mass of the β particle is exceedingly minute in comparison.

By an extremely ingenious device, whereby the number of particles which enter a tiny hole at a distance from a quantity of radium could be counted, Rutherford has been able to determine that the number of α particles which are projected from a single gram of radium in equilibrium with its products, reaches the enormous number of 13.6×10^{10}, or 136,000,000,000 *per second;* and yet, with this process continuing uninterruptedly through the centuries, a gram of radium would lose only one-half of its mass after the lapse of 2000 years.

The energy of the α particles, as they are thus projected from the atoms of radium, is likewise inconceiveably great, being no less than six hundred million times as great as the energy of a rifle bullet, mass for mass. So great, in truth, is the energy of these tiny particles, that their impact upon the mass of the radium itself maintains this substance constantly at a temperature higher than that of surrounding objects. Indeed, from the energy manifested by the α particles, it has been computed that a single gram of radium, distributed through each cubic yard of the sun's material, would fully account for the entire amount of energy in the form of light and heat which this luminary sends out.

531. The Disintegration Products of Radium. After the radium atom has projected an α particle, together with a slowly-moving β particle, the remaining portion ceases to be radium, and becomes a heavy gas, called the emanation. The emanation is likewise radio-active and projects an α particle from its substance, after which it becomes once more a solid that is deposited upon surrounding objects, and is called radium A. Radium A also throws out an α particle, and passes over to an entirely different substance, known as radium B. This substance then undergoes a change, whereby a β particle is thrown out and γ rays produced, and then it is transformed into a still different kind of matter, known as Radium C. This element then loses an α and a β particle, the latter, as always, accompanied by the γ rays, after which it becomes a form of matter, known as radium D. From radium D is projected a β particle but moving so slowly that γ rays are not produced, and an entirely different substance, known as radium E, is the result. Radium E then loses a swiftly-moving β particle, with its attendant γ rays, and passes over to a form of matter known as radium F, which is identical with the element polonium, previously isolated by Mme. Curie. This substance loses an α particle and leaves a body which is not radio-active but apparently stable, and which may be called radium G.

The successive changes which thus take place in an atom of radium and its resultant products may be illustrated in Fig. 346. From this diagram it will be observed that the radium

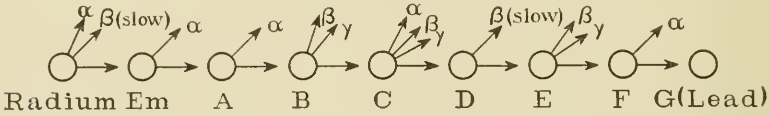

FIG. 346. The disintegration products of radium.

atom, which has an atomic weight of 226.5, after losing five α particles, each of which has an atomic weight of 4, leaves, as a residue, radium G, which is not radio-active, and which has an atomic weight of 206.5. This, however, is the atomic weight of lead; and it is therefore believed that what we call common lead is the final disintegration product of radium. Furthermore there is evidence to show that radium itself, in all probability, is a descendant of the element uranium, three generations of radium products having intervened between the two.

532. The Transmutation of the Elements. In order to test experimentally the conclusion that the α particle is really the helium atom, as seemed probable from its atomic weight, Sir William Ramsay enclosed in a glass tube a quantity of the emanation of radium and observed its spectrum. In a trifle less than four days, which is the period of the transformation of this gas, and during which the α particles which escaped from the emanation had collected in large numbers in the tube, he examined its spectrum again, and found, indeed, that the spectrum of radium emanation had changed completely to that of helium, showing conclusively that the α particle is nothing else than the positively charged helium atom.

It is thus demonstrated that, by losing a portion of itself, an atom of matter can pass into the form of another substance, and the theory of the transmutation of the elements may therefore be regarded as an established fact. Moreover, since all elements and all substances are found to be in a greater or less degree radio-active, it is possible that all known substances are slowly changing into substances of a different nature.

533. The Possible Structure of the Atom. In order to account for the various phenomena of radio-activity, and to offer an explanation of the enormous store of energy resident within an atom, Thomson has conceived of an atomic structure which, in part at least, serves to interpret the facts deduced from observation and experiment. According to his original view, an atom consists of a number of electrons revolving in a medium of "positive electrification," the number of the electrons being probably three times as great as the atomic weight of the element. These electrons were conceived to revolve in orbits, the number of electrons in each orbit being proportional to the distance of the orbit from the center of the atom, while the orbits themselves were arranged in concentric rings, and perhaps in parallel planes. When the electrons are displaced for any reason, they vibrate back and forth about their positions of equilibrium, thereby sending out electromagnetic waves, to which the phenomenon of radiation is believed to be due. The stability of an atom depends upon the number of electrons in the various concentric rings, an extremely stable system being able to take on one more electron, while an unstable ring is liable to lose one. This difference in the stability of the rings may be supposed to account for the various degrees of activity of the different chemical elements.

This view was later modified by assuming that the atom consists of a complex structure of positively and negatively charged particles revolving with enormous velocities within the atom, and held in equilibrium by their mutual attractions. From the scattering of the α particles, produced when these particles pass through the atoms of matter, Rutherford, however, has concluded that the atom consists of a concentrated central positive charge, about which the electrons revolve in orbits, as the planets revolve about the sun. The positive central nucleus, to which the larger part of the mass of the atom is due, is thought to consist of a group of positively charged helium atoms, or possibly of hydrogen and helium atoms, each helium atom having lost two electrons, and therefore charged positively with a double ionic charge. The positively charged helium atoms at the center and the negatively charged electrons revolving about them must form a system that is electrically neutral, and normally held in equilibrium by the mutual attraction of these oppositely charged particles.

Possibly because of the loss of energy produced by the radiation which is constantly taking place from the atom, or for some other reason which at present is not understood, the central system becomes unstable, and a helium atom escapes. As it leaves this system, it must acquire a large part of its energy in moving through the intense field produced by the positively charged nucleus. This escape of an α particle must alter the system of forces holding the electrons in equilibrium, so that ultimately the electrons in a particular orbit become unstable, and one of their number escapes as a β particle. It is possible that the velocity of this particle in its orbit is decreased as it leaves the atom, due to the attraction exerted by the strong central field.

If this theory is correct, the enormous energy which the α and the β particles manifest at the time of their escape from the atom must be obtained from the energy, either potential or kinetic, which these particles possessed within the atom before being expelled. Moreover, it is evident not only that the electrons are a component part of the atoms of all kinds of matter, but also that the atoms of helium, and probably also the atoms of hydrogen as well, are the elementary substances out of which the atoms of other elements are built up.

We may therefore possibly form the conception of an atom of radium by imagining this atom to be a particle of this substance no larger than the one thousand million millionth of the point of the finest cambric needle. At the center of this atom is a complex structure of positively charged helium atoms, which are themselves in rapid rotation according to definite laws, and around this central nucleus is a swarm of electrons each about as large, in comparison with the size of the atom, as a grain of sand is to a church, all revolving in tiny orbits at the rate of one hundred and eighty thousand miles in a second of time, and making five hundred million million revolutions per second.

SOUND

CHAPTER XLVI.

WAVE MOTION.

534. Vibrations. Sound may be regarded either as a subjective mental impression or as an objective material occurrence. The subjective consideration deals with the construction and arrangement of the parts of the ear, and with the processes by which a vibration after reaching this organ is conveyed to the brain, thereby producing the sensation of sound. The objective treatment inquires as to the conditions of the sounding body as well as to the motions which are there taking place, and investigates the nature and characteristics of the aerial vibrations, by means of which a disturbance is transmitted to the ear. It is mainly when sound is considered in this latter or objective sense, that is, when it is regarded as a purely physical phenomenon, that we shall undertake the study of this subject.

The first important observation to be made in the science of acoustics is the fact that every sounding body is in a state of vibration. It is the vibration of the body that sets in vibration the surrounding air, and these latter vibrations are propagated from particle to particle through the air until they strike the ear. That the air, in truth, is the medium by which, in general, the vibrations of a sounding body are transmitted to the ear may be shown by placing a bell under the receiver of an air pump. At first the bell sounds loudly; but as the exhaustion of the air progresses, the sound becomes fainter and fainter, until at last it is hardly perceptible. No sound whatever can travel through a perfect vacuum.

Illustrations of the fact that a sounding body is itself in vibration abound on every hand. The prongs of a tuning fork are observed to be in rapid motion whenever it emits a note. Indeed, the fork can be made to give forth a sound only when a force, such as a blow or a rosined bow, acts upon the prongs to distort them, after which the great elasticity of the steel causes the vibration to be maintained for some time. A little ball hung by a thread against one of the vibrating prongs, Fig. 347, will be violently driven away by the blows. A stretched string becomes a source of sound when it has been drawn to one side and released, so that the string is thrown into a state of vibration. If a little paper rider is placed upon the string, it will be thrown off at once whenever the string is sounded, showing that the reproduction of sound depends primarily upon the vibration of the sounding body.

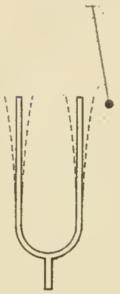

Fig. 347. A vibrating tuning fork.

A simple experiment illustrates the important fact, which must constantly be borne in mind, that the sounding body itself is a source of energy. As a tuning fork is made to vibrate in the air, the prongs have only the air to set in motion, hence the sound persists for some time. If, however, the prongs are dipped beneath the surface of water, the latter will be spattered out in numberless small drops, and the sound produced by the fork quickly ceases. The greater mass and the consequently greater inertia of the water causes an amount of work to be done in moving this medium greater than that performed than in setting in motion the lighter air, hence the energy represented by the vibrating fork is more rapidly exhausted, and the fork comes more quickly to rest.

535. The Vibrations Take Place with Harmonic Motion. When a body is distorted, elastic forces are instantly called into play, and the force of restitution (p. 114) tends to make the deformed body resume its original condition. A study of the laws of elasticity develops the fact that the force producing the distortion, and therefore also the force of restitution, is directly proportional to the amount of distortion, or to the displacement. Accordingly, we have a body moving under the action of a force which varies directly as the displacement, and this, as we have learned (p. 71), is the characteristic of harmonic motion.

Every particle of a sounding body, therefore, vibrates back and forth in its own little path with harmonic motion. When the elastic body is distorted and the distorting force is removed, the force of restitution tends to restore the body to its original undeformed state. The inertia of the particles of the body, however, causes them to swing past their state of equilibrium until they are distorted in the opposite direction, from which it follows that the same phenomenon takes place with each particle of a sounding body that takes place in the swinging of a pendulum.

It is a simple and beautiful experiment to show that the prongs of a tuning fork, for instance, really do vibrate with harmonic motion. If one of the prongs is provided with a stylus. and the fork then set in vibration, the stylus will trace a wavy line as it is drawn uniformly over a piece of smoked glass. This wavy line is at once recognized as the curve of sines, or the harmonic curve, which we shall see later can be produced only by the combination of uniform motion and harmonic motion, the two motions taking place at right angles to each other. From this it follows that, if the fork is drawn along with uniform motion and the resulting curve is the harmonic curve, the motion of the stylus, and therefore that of the prong, must be harmonic.

536. The Manometric Flame. When an organ pipe is sounded, it may be shown that the entire column of air is thrown into vibration, the movements being parallel to the length of the pipe. If a sheet of paper is placed over the top of an open pipe, the vibrations of the air cause a fluttering of the paper. Moreover, if a piece of paper, held in a horizontal position by strings, is sprinkled with sand and lowered into the interior of the pipe, the rattling of the sand upon the paper shows that the latter is thrown into violent vibrations.

Far better than the above method of studying the vibrations of an air column, is the device known as the manometric flame. Over a hole in the side of the tube containing the vibrating column of air, is stretched a thin rubber diaphragm, which, in turn, is covered by a little circular box, Fig. 348. Only the rubber diaphragm, therefore, separates the interior of the tube from the cavity of the little box, or capsule. Two tubes are inserted in this box, through one of which illuminating gas

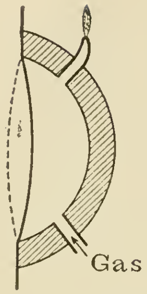

FIG. 348. The manometric capsule.

enters this cavity, while from the other it escapes and may be lighted as a little jet. When the air within the tube or organ pipe is at rest, the rubber diaphragm is not in any way distended, and the little jet of gas burns quietly. As soon, however, as the air within the tube is made to vibrate, the rubber diaphragm also is thrown into vibration so that it swings back and forth, alternately compressing and rarefying the gas in the interior of the capsule. This variation in the pressure of the gas in this manometric capsule causes the little flame to flicker, burning higher when the gas within the capsule is compressed and lower when it is rarefied.

Since the vibrations of the flame take place with great rapidity, the eye cannot follow them directly; but by carrying the head rapidly from side to side, the observer will perceive a succession of images of the flame, as these are formed in different positions upon the retina. The same effect may be accomplished by using a revolving mirror, which, as it turns, shows the image of the flame in different positions. If the flame is steady, its image will simply be drawn out into a broad band of light, and this will be observed in the mirror; but if it vibrates, these successive images will be alternately long and short, so that the vibrating flame, as seen in the mirror, is serrated, Fig. 349, resembling saw-teeth. The small inertia of the rubber membrane causes it faithfully to follow the vibrations of the air column upon which it is placed, so that this has proved to be one of the most valuable methods for studying aerial vibrations.

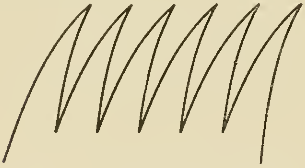

FIG. 349. The manometric flame, as seen in a revolving mirror.

537. The Phonograph. The phonograph, as devised by Edison, is a beautiful illustration of the principle that a sounding body is in a state of vibration. It demonstrates, also, the fact that when a body is thrown into suitable vibrations, a sound will be produced. This instrument consists in its essential features of a diaphragm, not unlike that of a telephone, but provided with a little stylus. Immediately beneath this stylus a wax cylinder turns with great uniformity of speed, and moves slowly along in the direction of its length, being thus carried by a very fine screw. The instrument is so adjusted that the stylus just touches the cylinder as the latter turns, thereby making a very faint spiral line upon the wax surface. If the diaphragm advances, the stylus also advances, and makes a deeper indentation in the wax; and conversely, if the diapragm recedes, the stylus likeswise recedes and makes a shallower impression. Accordingly, when speech is uttered before the instrument, the vibrations of the air throw the diaphragm into a series of vibrations similar to those of the air which produce the sound, and these vibrations are thus recorded as so many indentations in the surface of the wax, Fig. 350.

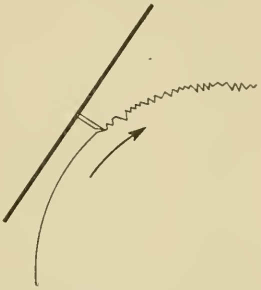

FIG. 350. The record on a phonograph cylinder.

If the cylinder is now turned back to the starting point, and the little stylus replaced by a tiny sapphire ball, this minute sphere will be moved in and out as the indentations in the waxed cylinder are carried along under it. The movements of this little ball are communicated to the diaphragm, and this is therefore caused again to execute vibrations in every way precisely similar to those which it originally made when the sound was produced before it. Accordingly, whatever words were spoken before the diaphram when the record was made upon the wax, will now be reproduced when the little sapphire ball moves over this record, since the vibrations of the diaphragm are the same in both cases.

538. The Production of Wave Motion. When the particles of a vibrating body move back and forth, they must necessarily set in motion the particles of the medium in which the body is placed. The particles of the medium, therefore, are thrown into vibrations which are similar in every way to those which the body itself executes. If, then, the medium in which the vibrating body is placed is elastic, its particles, when moved, tend to displace others, and these still others, and so on, and hence a disturbance which is started in an elastic medium travels forward in that medium. Moreover, if the original vibrating body executes regular vibrations, this motion is communicated to each particle of the medium in which the body is placed, and hence, after the lapse of a certain time, each particle of the medium executes regular vibrations similar in every way to those of the body. This communication of the disturbance from particle to particle throughout the medium is called the *propagation* of the disturbance. A careful study must be made of the means by which this disturbance is propagated.

539. Progressive Motion of the Wave. We will, accordingly, assume that we have an elastic medium, one particle of which is forced to execute regular vibrations with harmonic motion. We will also assume that this vibrating particle is the one marked 1, Fig. 351, and that it is forced to execute harmonic motion in the vertical direction. If, for the sake of convenience, we imagine the period of the vibration to be divided into twelve equal intervals, we can locate the position of this particle at intervals of 1/12 of a complete vibration. Starting then from *O* as the position of equilibrium, the vibrating particle occupies successively the positions, *A*, *B*, *C*, *B*, *A*, *O*, -*A*, -*B*, -*C*, -*B*, -*A*, *O* at intervals of 1/12 of a vibration. We will further assume that the particles of the medium which are about to be disturbed may be represented by the horizontal row of numbered black dots, and that each of these dots is bound to its neighbor by a perfectly elastic medium. Moreover, these particles are to be considered as being so close together that their movements take place parallel to one another. The letters *A*, *B*, *C*, etc., represent, therefore, not only the various positions of particle 1, but also the successive positions which each of the other particles occupies as it is compelled also to execute vibrations in the vertical direction.

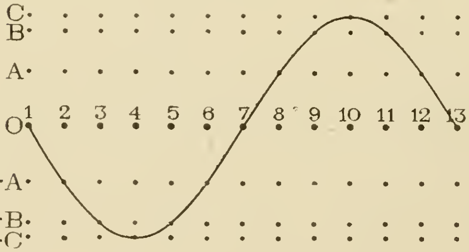

FIG. 351. The propagation of a transverse wave.

Accordingly, when particle 1 moves to the position *A*, the medium between 1 and 2 is distended, the inertia of 2 is being overcome, and 2 is about to move upward. When 1 reaches the position *B*, 2 has reached the position *A*, the medium between 2 and 3 is stretched, and 3 is about to move upward. When 1 reaches the position *C*, 2 has reached the position *B*, 3 has reached the position *A*, and 4 is about to move. Particle 1, now having reached its maximum displacement, moves back towards its position of equilibrium, reaching *B* just as 2 reaches *C*, 3 reaches *B*, 4 reaches *A*, and 5 is about to start. When 1 reaches *A*, 2 has arrived at *B* on its backward path, 3 has attained its greatest elongation at *C*, 4 has reached *B*, 5 has come to the position *A*, and 6 is about to start. In the same way, 1 comes to *O* when 7 is just about to start, to -*A* when 8 is about to move, to -*B* when 9 is upon the point of moving, and to -*C* when 10 is beginning to move. 1, having reached its maximum negative elongation, now moves back to -*B* just as 2 reaches its maximum negative elongation at -*C*, 3 reaches -*B*, etc., and 11 is about to move. So, also, when 1 comes back to -*A*, 12 is about to move, and when 1 reaches *O*, having completed one whole vibration, the disturbance has just reached 13, and this particle is now about to move.

The continuous line drawn through the positions which the various particles occupy at any instant, gives a curve which is called a *wave*, or the *curve of sines*, or the *harmonic curve*. Since at any instant the disturbance among the particles of the medium exists in the form of a wave, the motion of this disturbance through the medium is called wave motion. Accordingly, while the wave has a forward movement, the particles themselves move simply up and down, so that it is the wave motion alone that is propagated through the medium.

540. The Wave Length—Transverse Waves. By following out in detail this process as here given, it is seen that each particle moves only after the particle immediately before it has moved, and that each particle is executing a vibration precisely similar to the one immediately preceding it. We can, therefore, establish the general rule that when a vibration occurs in an elastic medium, every particle of that medium is thrown into a vibration precisely similar to that of the first particle, but that the vibration of each particle occurs a little later than that of the one immediately preceding. Thus, by referring again to Fig. 351, we see that each particle of the medium between 1 and 13 is executing harmonic motion precisely similar in period and amplitude to the first particle, but that it is in a *different phase of vibration* (p. 59) from every other particle, 2 being 1/12 behind 1, 3 being 1/12 behind 2 and 1/6 behind 1, and so on. From this it follows that while every particle of the medium vibrates, it vibrates *successively*, being moved by the particle immediately before it, and moving, in turn, the particle immediately behind it. The disturbance, therefore, does not travel instantaneously, but progress with a definite velocity which depends upon the elasticity of the medium and upon the inertia of its particles, that is, upon the density. Accordingly, the disturbance has travelled a perfectly definite distance along the medium during the time of one complete vibration of the first particle, and this distance is known as the wave length. *A wave length, therefore, is the distance which the disturbance has travelled during the time of one complete vibration of the first particle.* This distance is usually denoted by the Greek letter λ (lambda).

But particle 1, after completing its first vibration, is about to move upward again, thus beginning its second. As regards its state of motion, therefore, it is in precisely the condition of 13, for this particle also is about to move upward, thus beginning its first vibration. 1 and 13 are, then, in exactly the same phase of motion, 1 having completed one entire vibration and being now about to begin the second, and 13 having until this this moment remained at rest, and being now about to begin its first vibration. As we have seen above, however, 1 and 13 are a wave length apart. *A wave length*, therefore, may be defined as *the distance between two consecutive particles which are executing similar vibrations at the same time*, or as *the distance between two consecutive particles which are in the same phase of vibration.*

A wave, such as the one here described, in which *the direction of propagation is at right angles to the direction of vibration of the particles*, is called a *transverse* wave. The *amplitude* of the vibration is the greatest distance which the vibrating particle attains from the position of equilibrium, that is, it is the maximum displacement. The amplitude of the wave, therefore, is the same as the amplitude of the vibration. Inasmuch as the period of a vibrating particle is independent of the amplitude (p. 75), *the wave length of the disturbance in any medium is also independent of the amplitude.*

541. Equation of the Harmonic Curve. The results which we have here obtained graphically may also be derived algebraically. In expressing this relation, we must remember that the harmonic curve is a resultant motion, produced by harmonic motion taking place in the vertical direction and uniform motion in the horizontal direction. We may think of this curve, therefore, as being traced by the particle moving to the right with uniform motion while at the same time it is vibrating vertically with harmonic motion. The vertical distances, accordingly, represent the various displacements of the particle vibrating with harmonic motion, while the horizontal distances represent the distances which the wave has travelled with its uniform velocity of propagation. We have seen (p. 62) that the expression for the displacement x of a particle moving with harmonic motion is

$$x = A \cos \frac{2 \pi t}{T},$$

where A is the amplitude, T the period, and t the time which has elapsed since the particle left its position of equilibrium.

In the present instance, however, the displacement is in the y, or vertical, direction; hence the equation for the motion of a particle moving with harmonic motion in the vertical direction is

$$y = A \sin \frac{2 \pi t}{T}.$$

The wave moves forward with the velocity v in the horizontal, or x, direction, and therefore in the time t it has progressed with uniform motion the distance

$$x = vt.$$

Eliminating t between these two expressions,

$$y = A \sin \frac{2 \pi x}{v T}.$$

But v is the velocity with which the motion advances through the medium, that is, it is the distance which the disturbance travels in one second, and T is the period, or the time of one complete vibration of the particle. Accordingly, vT is the distance which the disturbance has travelled during one vibration of the first particle, and this distance, therefore, is the wave length. Letting λ represent the wave length, we have

$$\lambda = vT,$$

and the equation of the harmonic curve becomes

$$y = A \sin \frac{2 \pi x}{\lambda}.$$

Since the value of y depends upon the sine of an angle, it is evident that y is a *periodic* function, or, in other words, that the values of y recur *periodically*. Moreover, as the only variable in this expression is x, which represents the distance the wave has travelled, it follows that by giving to x different values, that is, by assuming different distances in the horizontal direction, we shall obtain different values for y, or the height of the corresponding point of the curve above or below the horizontal axis. When

$x = 0, \quad y = 0$
$x = \frac{1}{4} \lambda, \quad y = A$
$x = \frac{1}{2} \lambda, \quad y = 0$
$x = \frac{3}{4} \lambda, \quad y = -A$
$x = \lambda \quad y = 0.$

By plotting these values of x and y, the harmonic curve is obtained as before.

542. Longitudinal Waves. There is, however, a form of wave motion which differs from the one just described, in that the direction in which the wave is propagated lies *in* the direction in which the vibrations of the particles occur. Thus, as before, we can imagine the particles of the elastic medium to be represented by the row of black dots, Fig. 352, these being equidistant when the medium is undisturbed. If we now imagine the first particle to be displaced to the right, the medium between the first and second is compressed, and as this tends to regain its original condition, it pushes the second particle also to the right. This, in turn, compresses the medium between the second and third, so that the third particle moves likewise to the right, compressing the medium between the third and fourth, and crowding the fourth particle to the right. A disturbance of the nature of a compression, therefore, travels along the medium, each particle being compressed by the particle immediately preceding, and compressing in its turn the particle immediately following. Thus a condensation is propagated throughout the medium, the particles suffering compression successively.

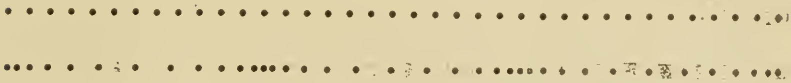

FIG. 352. A longitudinal wave consists of condensations and wave factions.

But we have assumed that we are dealing with an *elastic* medium, so that the first particle, after it has been displaced, vibrates back and forth with harmonic motion. Having, therefore, reached its maximum displacement to the right, this particle starts to move backwards towards its position of equilibrium, but in so doing it acquires a velocity which carries it through this point to an equal displacement in the opposite direction, in this case to the left. Accordingly, as the first particle begins to move backward, the elastic medium between the first and second particles is distended, and as the first particle swings through its position of equilibrium with its maximum velocity, the medium between the first two particles is so widened that the distance between these particles is greater than that between them when the medium is undisturbed. This is expressed by saying that a rarefaction exists between these particles. This distended, or stretched, medium, produced by the first particle moving to the left, acts upon the second particle, moving it likewise to the left, and distending the medium between the second and third particles. The third, as it moves to the left, stretches the medium connecting it to the fourth particle, causing this, in its turn, also to move backward, and so on. Therefore, a motion backward is communicated from particle to particle, so that each particle in succession moves to the left, producing a wave of rarefaction which is propagated to the right through the medium.

When the first particle has reached its maximum elongation to the left, it at once begins its movement again to the right, thus starting a second wave of compression along the medium. The first compression or condensation reaches a perfectly definite distance when the second condensation starts, that is, it travels a perfectly definite distance during the time of one complete vibration of the first particle. This distance, therefore, is a wave length, and depends upon the voluminal elasticity of the medium and also upon the mass, that is, upon the density of the particles. Between these two particles, which thus mark the wave length, there are particles in every phase of vibration, and therefore, as before, the second definition of wave length applies also here: namely, *the wave length is the distance between two consecutive particles in the same phase of vibration.*

This form of vibration is known as a *longitudinal* vibration, and the wave which is thus produced, consisting of a condensation followed by a rarefaction, is known as a *longitudinal wave.* A *longitudinal* wave, therefore, is a wave in which *the direction of propagation is parallel to the direction of motion of the particles.* In longitudinal waves, as in transverse waves, the wave length is independent of the amplitude of the vibrating particles.

In a *transverse* wave the medium is thrown into a series of *crests and troughs*, the distance from crest to crest or from trough to trough being a wave length. In a *longitudinal* wave, the medium is divided into *condensations and rarefactions,* the distance from condensation to condensation or from rarefaction to rarefaction being a wave length. Whether the wave which is propagated from a disturbance in a medium is a transverse or a longitudinal wave depends upon the nature of the elasticity of the medium. *If the medium has elasticity of volume but not of form*, as is the case with liquids or gases, *only longitudinal vibrations can be propagated.* The vibration of waves of sound in the air, therefore, are wholly longitudinal. *If, however, the medium has also elasticity of form*, as is the case with an elastic solid, *transverse as well as longitudinal waves* can be propagated through the medium. The vibrations of the ether which produce light are known to be transverse, and hence the ether is believed to be an elastic solid. The longitudinal waves in the ether, if, indeed, these exist at all, have never been detected.

543. Spherical Waves. When a vibration occurs in a homogeneous medium, that is, a medium which is alike on all sides, the disturbance is propagated not only in one direction but in all directions. At the end of a certain time, therefore, the disturbance which starts from the vibrating particle has been propagated an equal distance in all directions, and is found upon the surface of an ever-enlarging sphere which has the vibrating body as a center. The rate at which the radius of this sphere increases is the velocity of propagation of the wave motion in the medium. In a homogeneous medium, therefore, the disturbance is propagated throughout that medium in the form of a spherical wave. Moreover, every particle of the medium must execute vibrations precisely similar to those of the source of disturbance at the center, and every particle on the surface of this spherical wave must be in the same phase of vibration. *This surface of the spherical wave, of which the vibrating body is the center, is called the wave front.* The *wave front*, therefore, may be defined as *the continuous locus of all particles which are in the same phase of vibration.* If the source of disturbance is very far away, so that the radius of this sphere is very great, a small portion of this spherical wave is practically a plane, and is spoken of as a plane wave, or a plane wave front.

CHAPTER XLVII.

REFLECTION OF WAVE MOTION.

544. Reflection of Wave Motion. The reflection of a wave means the turning back of the wave. When reflection occurs, the wave ceases to advance in the direction in which it has been moving and is turned back, so that it again traverses the medium over which it has travelled, this time moving in the opposite direction. *Whenever a wave meets a medium in any way different from the one in which it has been travelling, reflection occurs.*

545. Reflection of Longitudinal Waves. The reflection of a longitudinal wave may be illustrated as follows: If we imagine a number of ivory balls, five for example, to be suspended horizontally in a row, these will represent the individual particles of an elastic medium. If, now, the first is drawn back and allowed to fall upon the second, Fig. 353, the force of the impact slightly flattens this ball. As this regains its original shape, it strikes an equal blow upon the third, flattening it, and so on. In this way a wave of compression is propagated through the row, in this case moving to the right, each ball being distorted by the one immediately preceding, and in turn distorting the one immediately following. If, now, a block of iron or other immovable obstacle is held rigidly in position, just touching the

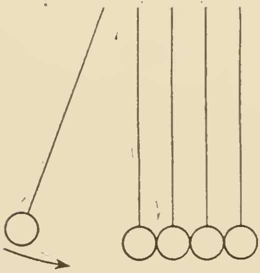

Fig. 353. Transmission of a longitudinal wave.

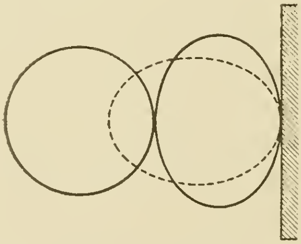

Fig. 354. Reflection of a longitudinal wave with change of sign.

last ball, Fig. 354, this fifth ball, when flattened, is unable to move the iron, and hence strikes the fourth ball a blow as it regains its original condition. The blow which is thus delivered tends to move the fourth ball to the left; but as this cannot move, due to the presence of the third ball, it becomes flattened, and as it regains its spherical condition it strikes the third ball a blow. In the same way, the third ball distorts the second, and the second delivers a blow upon the first, forcing it to fly out to the left. The disturbance, which was first propagated through the medium moving to the right, has thus travelled a second time over the medium, this time in the reverse direction, and hence the wave has been reflected at the point where the ivory medium ceased and the iron began.

It is to be observed that each ivory ball was successively compressed as the reflected wave travelled through it, and hence in this instance a wave of compression is reflected as a wave of compression. It is, however, also to be observed that the first particle moved to the right as the direct wave advanced, and that it now moves to the left as the reflected wave returns. If motion to the *right*, therefore, is regarded as *positive*, motion to the *left* is considered *negative*, and *the sign of the motion of the first particle has been changed by the reflection of the wave.* The sign of the motion of the first particle of the medium, as it moves under the action of the reflected wave, gives name to the reflection which has taken place. In the present instance the direction of the motion of the first particle has been changed, hence reflection occurred *with change of sign.* Accordingly, reflection *with change of sign* takes place whenever a wave in any medium meets a medium *more dense*, that is, one in which it can *move with less freedom*, than the one in which it has been travelling. When a *longitudinal* wave is reflected *with* change of sign, a *condensation* is reflected as a *condensation*, and in the same way a *rarefaction* is reflected as a *rarefaction.*

If we now imagine the iron block removed from the last ivory ball, a wave of compression which is now, as before, sent to the right, causes the last ball to fly out, Fig. 355, moving much farther than any of the intermediate balls. If we now think of each ball as connected with its neighbor by an elastic cord, the cord will be stretched when the last ball moves, and this will drag the fourth one after it, moving also to the right. As this ball moves it will, in the same way, drag the third one after it, this will drag the second likewise to the right, and this will cause the first to move also to the right. The disturbance has therefore been propagated backward through the medium, and accordingly reflection must have occurred at the point where the row of ivory balls ceased. The motion of the first particle, and indeed, that of every particle in the reflected wave, is *in the same direction as in the direct wave.* The sign of motion of the first particle has not been changed, and therefore *reflection has taken place without change of sign.*

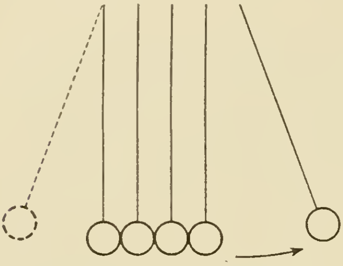

FIG. 355. Reflection of a longitudinal wave without change of sign.

Reflection without change of sign occurs whenever a wave in any medium meets a medium *less dense*, that is, one in which it can *move with greater freedom* than the one in which it has been travelling. It is to be noticed that the fifth particle moves much more freely in the air than in the ivory medium; and as it moves it distends the medium connecting it with the fourth particle, and so on, hence the *compression* which passes down the medium is, in this instance, reflected as a *rarefaction*. That is, the last particle in the wave of condensation which passes down the medium changes at once, as soon as it moves, to the first particle in the wave of rarefaction which passes back over the medium. Therefore, when a *longitudinal* wave is reflected *without* change of sign, a *condensation* is reflected as a *rarefaction*, and, conversely, a *rarefaction* is reflected as a *condensation*.

In this connection one very important matter is most carefully to be noted. In a *longitudinal* wave, when reflection occurs *with change of sign, there is no change of phase at the point of reflection*, a condensation which traverses the medium in the medium in the direct wave being reflected as a condensation in the reflected wave. When, however, in a *longitudinal* wave, reflection *without* change of sign takes place, a *condensation* is reflected as a *rarefaction*, that is, the *disturbance changes immediately to the opposite phase, and a loss of a half wave length occurs at the point of reflection.*

546. Reflection of Transverse Waves. A similar analysis of reflection may be made with transverse waves. If we imagine the medium in this case to be a stretched rope or a spiral spring, extending from the table to the ceiling, Fig. 356, a displacement of one particle to the right will drag the next particle in the same direction, and hence a crest of displacements to the right will pass up the spring. At the top, the displacement of the particles to the right exerts a force upon the ceiling tending to move it also to the right; but as this cannot move, the reaction of the ceiling upon the spring urges the particles of the latter to move to the left. As these move to the left, they drag with them the particles immediately below, and hence a wave in which the particles move to the left travels down the spring. The wave, therefore, has been reflected where the medium changes at the ceiling. Moreover, when the reflected wave reaches the bottom of the spring, the motion of the first particle is to the left, or in the direction opposite to its motion in the direct wave, and therefore at the ceiling the reflection occurred *with change of sign*. A *crest* which is sent up the spring is reflected as a *trough*, and conversely a *trough* which is sent up is reflected as a *crest*. The uppermost particle of the spring at one instant is a portion of the ascending crest, and the next instant it is a portion of the descending trough. Accordingly, *in transverse waves, when reflection occurs with change of sign, there is a loss of half a wave length at the point of reflection.*

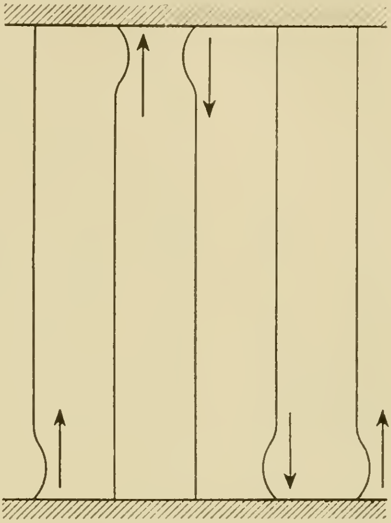

FIG. 356. Reflection of a transverse wave with change of sign.

To illustrate reflection without change of sign, we may imagine a heavy rope attached to the ceiling by a light cord, Fig. 357. If, now a disturbance to the right is sent up the rope, the particles move successively and equally to the right until the cord is reached. As the last particles of the rope are disturbed, they have only the particles of the lighter cord above to set in motion, and hence these last particles of the rope move out very much farther than any other particles. This increased motion of these particles disturbs the particles immediately beneath, and hence a reflected wave starts down the rope, being reflected at the point where the heavy rope is attached to the lighter cord. Moreover, the disturbance of the particles, which was to the right in the direct wave, is also to the right in the reflected wave, hence *reflection occurs without change of sign.* The *crest* which is sent up comes back as a *crest,* and, in the same way, a *trough* which is sent up comes back as a *trough.* Accordingly, in *transverse* waves when reflection *without* change of sign occurs, *there is no change of phase at the point of reflection.*

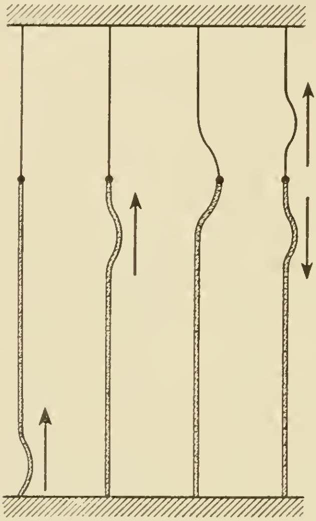

FIG. 357. Reflection of a transverse wave without change of sign.

The whole matter of the reflection of waves may be summed up as follows: *Whenever a wave, either longitudinal or transverse, meets a denser medium,* that is, a medium in which it can move with *less freedom* than the one in which it has been traveling, *reflection with change of sign occurs. Whenever a wave, either longitudinal or transverse, meets a rarer medium,* that is, a medium in which it can move with *greater freedom* than the one in which it has been travelling, *reflection without change of sign occurs.* In *transverse* waves when reflection *with* change of sign occurs, and in *longitudinal* waves when reflection *without* change of sign occurs, *there is a loss of half a wave length at the point of reflection.* The latter statement may be expressed by saying that *whenever the phase of the motion changes at the point of reflection a loss of half a wave length* *takes place.* This holds for transverse and for longitudinal waves alike.

547. Stationary Waves. On account of the difficulty of representing by a diagram the condensations and rarefactions of a longitudinal wave, a study of stationary waves may best be made by considering transverse waves.

Accordingly, we will once more consider that a disturbance in the form of a crest is given to a spiral spring, Fig. 358, which is stretched from the table to the ceiling. This crest is propagated to the top, is there reflected with change of sign, comes down as a trough, is reflected at the bottom, also with change of sign, and starts for the second time up the spring as a crest. If a second impluse in the same direction is given to the spring just as the first thus starts up for the second time, the two add themselves together and a crest of increased amplitude passes up the spring.

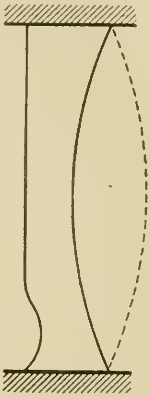

FIG. 358. The formation of a stationary wave.

If, now, the particles at the bottom are made to execute regular vibrations, the period of the vibration being the time required for the disturbance to travel up and down the spring, the impulses which are given by these vibrations occur in the same phase as the disturbance already upon the spring, so that the amplitude of the resulting disturbance is the sum of the amplitudes of the separate disturbances. Moreover, the period of vibration of all the particles is the same as that of the particles at the bottom; and as the disturbance which thus passes repeatedly up and down the spring gradually displaces more and more particles above and below itself, it soon follows that all the particles of the spring move simultaneously to the right as the crest passes up, and to the left as the trough passes down. The entire spring, therefore, vibrates as a whole, swinging first to the right and then to the left, and the progressive motion of the wave is altogether lost to view. This constitutes a *stationary* wave. *A stationary wave, therefore, is a wave in which the progressive motion cannot be detected, and is produced by a* *direct and a reflected wave existing upon the medium at the same time.*

If the stationary wave is *transverse,* there are always at least *two stationary points* upon the medium, called the *nodes.* In the present instance, the stationary points, or nodes, are at the top and bottom. The region between the nodes is called the *antinode,* and here the particles of the medium are all in the same phase of vibration. Since the disturbance, in the case of the spring, passes up and comes down during the time of one vibration of the first particle, *the wave length is twice the length of the spring,* or twice the distance between the nodes. In general, it may be said that *the wave length of any stationary wave is twice the distance between two adjacent nodes,* or, in other words, the distance between two adjacent nodes is half a wave length. When a medium such as a stretched cord or string vibrates as a whole, having nodes only at the two fixed ends, the medium is then thrown into a vibration having the longest wave length of which it is capable, and this is called the *fundamental vibration.* The note which is emitted by a sounding body vibrating with its fundamental vibration is called its *fundamental tone,* or, simply, *its fundamental.*

548. The Subdivision of the Vibrating Medium into Segments. Referring again to the long stretched vertical spring, we see that if the second disturbance is sent up just as the first has reached the top and is starting down, the crest of the second direct wave meets the trough of the first reflected wave at the center. This central particle is thus urged in opposite directions at the same time, and hence remains stationary. Accordingly, a node is determined at the center of the spring, Fig. 359, due to the fact that the direct and the reflected waves meet at this point in opposite phase. The spring is therefore divided into two vibrating segments with a node between them. In the formation of these subdivisions, and in the production of stationary waves in general, it is to be observed that the particles of the medium in adjacent segments are always in opposite phases of vibration.

Since, in the present instance, the wave length of this segmental vibration is just equal to the length of the spring, or twice the distance between two adjacent nodes, the note emitted by these vibrating segments is an octave higher than the fundamental tone, as will be shown later. The musical notes emitted by a medium which subdivides into an *exact whole number* of vibrating segments are known as the *overtones* or the *harmonics* of the fundamental. A string or cord which thus divides into two segments is then emitting its first overtone or harmonic.

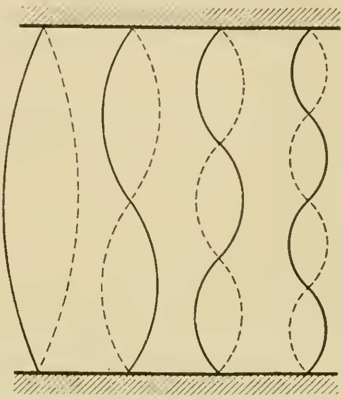

FIG. 359. The fundamental vibration of a cord, and its subdivision into two, three, and four segments.

If the second disturbance is sent up the spring when the first has travelled only two-thirds of the way to the top, the first reflected wave and the second direct wave meet in opposite phase at a point one-third of the length of the spring from the top, and thus determine a node at that point. A third direct wave which is now sent up meets the first reflected wave one-third of the distance from the bottom, and determines a node at this point as well. The medium is thus equally divided into three vibrating segments, and the note produced is the second overtone or harmonic of the fundamental.

In a similar way, if the first wave is half-way up the spring when the second one is started, the second direct wave meets the first reflected one in opposite phase at a point one-fourth of the length of the spring from the top, the third direct wave meets the first reflected one at the middle of the spring, and the fourth direct wave meets the first reflected one at a point one-fourth of the length of the spring from the bottom, determining a node at each of these points. The medium, therefore, is broken up into four vibrating segments, thus producing a note which is the third overtone or harmonic of the fundamental, and so on.

549. Transverse Vibrations of Cords. When a body like a stretched cord or string is made to vibrate *transversely*, it is the *tension* and not the elasticity of the cord that causes the vibration. Thus, a silk cord when tightly stretched, may be made to vibrate with transverse vibrations and to give forth a musical note, although the cord itself is very inelastic. When the stretched cord is displaced from its position of equilibrium, as, for instance, when it is bowed, it is a component of the tension and not the elasticity of the cord which causes the latter to return to its original position. This component of the tension, however, is proportional to the displacement, hence it takes the place of the force of restitution in deformed elastic bodies, and thus compels the cord, when released, to vibrate back and forth with harmonic motion. The velocity of propagation v of a disturbance on a stretched string depends, therefore, upon the tension t of the cord and upon its inertia, that is, upon the mass per unit length which the cord possesses. Letting v represent the velocity of propagation along the cord, t the tension, and d the mass per unit length of the cord, or the density, it may be shown that

$$v = \sqrt{\frac{t}{d}}.$$

Since the stretched cord is fastened at both ends, reflection with change of sign occurs whenever the disturbance which is propagated along the cord reaches the end. The co-existence of direct and reflected waves upon the cord throws the latter into stationary waves, and, as has been seen (p. 671), this can occur only when the cord is some whole number of half wave lengths long. Accordingly, if L represents the length of the cord, and if N is the number of segments into which the cord divides, λ being the wave length,

$$L = N\frac{\lambda}{2}. \qquad (1).$$

If

$$N = 1,$$

that is, if the cord is vibrating with but a single loop, the string is then executing its fundamental vibration, and

$$L = \frac{\lambda}{2},$$

or

$$\lambda = 2L.$$

If

$$N = 2.$$

that is, if there are two vibrating segments upon the string, the cord is emitting its first overtone, and

$$\lambda = L,$$

and so on. Therefore, by giving different values to N, we determine the wave lengths of the different segments into which the string may divide. The equivalent of this was done (p. 671) in the previous consideration of the subdivision of a string.

In the study of the propagation of wave motion, we have seen (p. 658) that the quantity known as the wave length, λ, is the distance which the disturbance has travelled during the time of one vibration of the first particle. If this particle is vibrating regularly, and is making n complete vibrations in a second, it is evident that during this second the disturbance in the medium will travel the distance $n\,\lambda$. But the distance which the disturbance travels in a second is the velocity, v, (p. 23), and hence

$$v = n\,\lambda.$$

Substituting this value of v in the expression

$$v = \sqrt{\frac{t}{d}},$$

we have

$$n\lambda = \sqrt{\frac{t}{d}}.$$

Replacing in this the value of λ, as derived from equation (1), we obtain, finally,

$$n = \frac{N}{2L}\sqrt{\frac{t}{d}}.$$

That is, the length, the mass per unit length of the cord, and the number of segments into which the cord divides remaining the same, *the number of vibrations per second* which a cord will make varies *directly* as the square root of the tension. Accordingly, if a cord is vibrating with its fundamental vibration under a certain tension, it will still vibrate with its fundamental vibration, but making twice as many vibrations per second, if the tension is increased four-fold. If we solve this last equation for N, the number of vibrating segments into which the cord divides,

$$N = 2Ln\sqrt{\frac{d}{t}}.$$

In other words, the length, the mass per unit length of the cord, and the number of vibrations per second remaining the same, *the number of segments* into which the cord divides varies *inversely* as the square root of the tension. If, therefore, under a certain tension a cord vibrates with its fundamental vibration, it will break up into two vibrating segments if the tension is made one-quarter of its former amount, and into three vibrating segments if the tension becomes one-ninth of its original amount, and so on.

550. The Transverse Vibrations of Rods. When a rod free at both ends is made to vibrate transversely, the fundamental vibration consists of three vibrating segments separated by two nodes, Fig. 360, the distance of the nodes from the ends being about one-quarter of the distance between them. Here, as in all other cases, the vibrations in adjacent segments are in opposite phase, so that while the two ends of the rod move upward, the central portion moves downward, and when the two ends move downward, the central portion moves upward. If the rod is bent into a U-shaped figure, the nodes approach each other, thus making the middle vibrating segments very much smaller. A rod so bent becomes a tuning fork. An examination of the vibrations of a tuning fork shows that the movements of the prongs take place in such a way that when one prong moves in the other also moves in, and when one moves out the other moves out. Moreover, it is likewise seen that the middle segment executes small vertical vibrations, so that the stem of the fork which is attached to this middle segment also moves with vertical vibrations. Thus, the vibrations of the fork are transmitted to the box, or resonator, upon which the fork is mounted, and in this way the sound of the fork is reenforced.

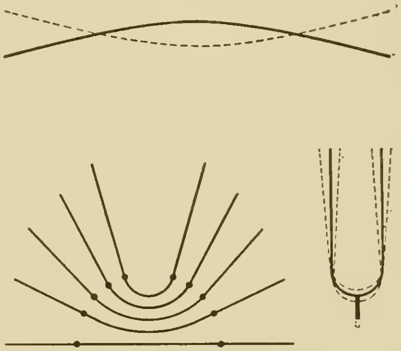

FIG. 360. The transverse vibrations of a rod and a tuning fork.

551. The Transverse Vibration of Plates—The Chladni Figures. A plate may be considered as a rod of very great width, and the vibrations of a plate may be regarded as those which would occur upon a broad rod. Chladni studied the vibrations of plates by sprinkling sand upon them, and observing the patterns formed as the sand was thrown away from the vibrating segments and collected along the lines of no vibration. Thus, Fig. 361, if we imagine the horizontal edges of the square brass plate free to vibrate, and the plate to vibrate with its fundamental vibration, it will divide itself, like a rod, into three vibrating sections, separated by two nodal lines. It is along these lines that the sand which may be scattered upon the plate will collect. If, now, the plate vibrates in more than one direction at the same time, the resulting vibration and the resulting sand pattern may be very complex. Wheatstone has shown that however complicated the vibration of a plate may be, the existing vibration may be considered as the resultant of two vibrations at right angles to each other which exist upon the plate at the same time. We have assumed that the horizontal edges of the plate vibrate, and have seen, Fig. 361, that the vibrating segments are separated by horizontal nodal lines. If the plate were to vibrate with its fundamental vibration and its vertical edges moving, it would be divided into three sections separated by two vertical nodal lines, as in Fig. 362.

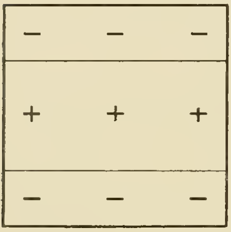

FIG. 361. A vibrating plate with horizontal nodal lines.

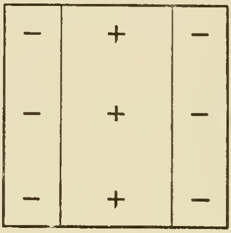

FIG. 362. A vibrating plate with vertical nodal lines.

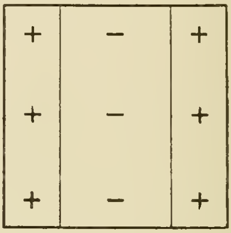

FIG. 363. Vibrations in phase opposite to those in Fig. 362.

If these two vibrations were now to exist upon the same plate at the same time, the principle of Wheatstone shows that the resulting vibration upon the plate will be the effect of compounding these two vibrations. In forming this resultant vibration, note must be taken of the phase of the second vibration, whether it is as shown in Fig. 362, or in Fig. 363. We may assume in these figures that the + sign indicates a movement of the portion of the plate towards the observer, while the — sign indicates a motion in the opposite direction. The superposition of the vibrations shown in Figs. 361 and 362 gives the resulting vibration as shown in Fig. 364, the central portion of the plate moving towards the observer and the corners of the plate moving away, while a diamond-shaped nodal line separates these two segments. If the vibrations represented in Figs. 361 and 363 are superposed, the resultant vibration is shown in Fig. 365, the nodal lines forming the diagonals of the plate. This latter pattern is one of the simplest of the Chladni figures to obtain.

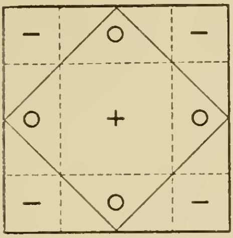

FIG. 364. Resultant vibration—Figs. 361 and 362 superposed.

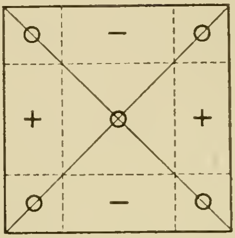

FIG. 365. Resultant vibration—Figs. 361 and 363 superposed.

By bowing the plate at different points, and stopping the vibration at others by touching these places with the finger, very complicated patterns indeed are formed by the sand as it is thrown away from the vibrating segments and collects along the nodal lines. Some of these designs are so complex that Wheatstone's analysis is not sufficient to resolve the resultant vibration of the plate into its component simple vibrations.

CHAPTER XLVIII.

THE VELOCITY OF SOUND.

552. The Velocity of Sound—The Direct Method. It is a matter of common observation that sound is not propagated instantaneously, and that a perfectly finite time is required for sounds to travel from one point to another. A workman at a distance is observed to deliver a blow with his hammer, but the report is not heard at that instant, due to the fact that time is required for the sound to travel across the intervening space. In the same way, if a gun is discharged, a distant observer by noting the puff of smoke, will recognize the fact almost instantly; the report, however, will be delayed, possibly several seconds, because it requires a certain time for the disturbance to be propagated this distance through the air.

Indeed, it was upon this principle that the earlier determinations of the velocity of sound were made. Two observers, each provided with a gun and also with a device for recording time, were stationed upon distant hills. The observer upon one hill discharged his gun, and the one upon the distant hill made careful note of the interval of time that elapsed between the perception of the flash and the sound of the report. The second observer then discharged his gun while the first made similar observations, and after repeating these experiments many times, the average of the whole series was taken. The distance between the two stations was then carefully measured; and as this distance and also the time required by the sound to travel this distance were then known, the velocity of sound could be approximately determined. It was found that sound travels a little faster with the wind than against it, and that shrill, acute sounds travel a little faster than deep, grave sounds. The pressure of the air has no effect upon the velocity of propagation, but the temperature has a wide influence upon the rate at which sound travels. When the average of a long series of observations is taken, it is found that the velocity of sound at 0° C. is very nearly 331 meters per second. The proximity of a stroke of lightning, therefore, may be computed by counting the seconds which elapse between the flash and the report.

Early in the nineteenth century many observations were made by a similar method upon the Lake of Geneva to determine the velocity of sound in water. A gun was discharged at the instant that a bell was struck under water, and an observer several miles distant listened for the sound of the bell by putting his ear to a trumpet, the mouth of which was underneath the surface of the water. By observing the interval elapsing between the flash of the gun and the sound of the bell as transmitted by the water, the velocity of sound in the water was determined. In this way the velocity of sound in water is found to be many times greater than in air; for although water is compressible only with great difficulty, it regains its volume completely as soon as the pressure is released, hence it is perfectly elastic. An experiment showing the rapid transmission of sound by solids may be easily tried. If an observer places his ear near an iron fence. and a second person some distance away strikes the fence a sharp blow, two reports will be heard, one coming through the fence, and the other, which is transmitted more slowly, travelling through the air.

553. The Velocity of Sound—The Indirect Method. This older method, which is known as the *direct* method of determining the velocity of sound, is open to many criticisms, principally due to the varying temperature of the different strata of air between the two stations, and especially to the fact that errors are inevitably made in recording the times at which the flash of the gun is seen and the report heard. For more accurate measurements of the velocity of sound, recourse is had to the *indirect* method, and this depends ultimately upon our definition of wave length.

We have seen (p. 658) that the wave length in an elastic medium is the distance which the disturbance or the wave travels during the time of one vibration of the first particle. If the elastic medium is the air, and if the vibrating particle is a portion of a sounding body, the length, λ, of the aerial wave is the distance through which the disturbance in the air has been propagated during the time of one complete vibration. Moreover, if the sounding body makes n vibrations per second, it is evident that $n \lambda$ is the distance which the disturbance travels through the air in one second, or the velocity v. Accordingly

$$v = n \lambda.$$

This equation may be called the fundamental equation in expressions involving the velocity of sound.

The process of computing the velocity of sound by the indirect method consists, therefore, in counting the number of vibrations per second which a sounding body is executing, and determining the wave length of this sound in the air. From the above reasoning, then, it is seen that the product of these two quantities gives the velocity directly. Measurements both of the number of vibrations per second and of the resultant wave length may be made with great accuracy, hence the velocity of sound may be determined with far less error by the indirect method than by the direct.

Another equivalent expression is often used instead of the form of the fundamental equation here given. If T is the time of vibration of the first particle, or the period, and n is the number of vibrations per second, it is evident that

$$T = \frac{1}{n},$$

and accordingly,

$$\lambda = vT.$$

This is the algebraic definition of the wave length, and is equivalent to the verbal definition given above.

554. Theoretical Determination of the Velocity of Sound. Whenever longitudinal vibrations are propagated in an elastic medium, it may be shown that two quantities only, both of which are properties of the medium, enter to affect the velocity of propagation, one of these being the elasticity E, and the other the density D. By the theory of elasticity it may be shown that the velocity of propagation v is connected with these quantities by the relation

$$v = \sqrt{\frac{E}{D}}$$

Inasmuch as the atmosphere is a medium having only voluminal elasticity and therefore transmitting only longitudinal waves, it was apparent to Newton that this relation must likewise express the velocity of sound, provided correct values for the elasticity and the density of the air were employed.

Newton reasoned that since sound is transmitted by a series of condensations and rarefactions, the air must be heated by the former and cooled by the latter; and as the amount of heating thus produced must just equal the cooling, he believed that the sound would travel through the air as if no change of temperature had occurred. It remained, accordingly, simply to replace E and D in this expression by the values of the isothermal elasticity (p. 310) and of the density of the air, respectively.

The isothermal elasticity of a gas was known to be numerically equal to the pressure at constant temperature (p. 312), and this, in the case of the free atmosphere, is 76 × 13.596 × 981 dynes per square centimeter (the modern notation is here used), while the density, or the mass of 1 cubic centimeter of the air at 0°C., is 0.001293 grams. When quantities corresponding to these were substituted in the above equation, that is,

$$v = \sqrt{\frac{E}{D}} = \sqrt{\frac{P}{D}} = \sqrt{\frac{76 \times 13.596 \times 981}{0.001293}} = 279.95$$

a value of 279.95 meters per second was obtained for the velocity *v*, which was only about 80 percent of the true velocity of sound, namely, 331 meters per second, as actually determined by experiment.

The cause of this great discrepancy was never explained until Laplace, in 1816, pointed out the error which obviously existed in Newton's computation. This investigator showed that while the air is warmed by the condensations and cooled by the rarefactions, the vibrations occur so rapidly that no exchange of heat can take place between these warmed and cooled portions and the surrounding air. Therefore, during the instant when the sound wave is passing, the air is in a condition in which it is compressed or expanded as if it were in a reservoir wholly impervious to heat, and accordingly in the adiabatic (p. 317) and not in the isothermal condition. Laplace then suggested that the adiabatic elasticity of the air instead of the isothermal elasticity should be used, for this would represent the condition of any small portion of the atmosphere at the instant when the sound wave was passing through it. Furthermore, by comparing Newton's computed value with the known velocity of sound as measured by the direct method, he showed that the adiabatic elasticity of the air must be greater than the isothermal in the ratio of 1.41 : 1, or that

$$v = \sqrt{\frac{E}{D} \times 1.41} = \sqrt{\frac{P}{D} \times 1.41} = 331 \text{ meters per second.}$$

From theoretical considerations he also proved that the ratio of the two elasticities of a gas must be the same as the ratio of the two specific heats (p. 318), and therefore that

$$\frac{E\text{ (adiabatic)}}{E\text{ (isothermal)}} = \frac{C_p}{C_v} = 1.41.$$

While modern measurements have confirmed the correctness of this deduction, it is undoubtedly true that, because of the difficulty in making accurate measurements of the two specific heats or of the two elasticities of a gas, the value of this ratio, as thus determined from the velocity of sound in the air, is more nearly correct than that determined by any other means.

The expression

$$v = \sqrt{\frac{E}{D}}$$

serves also to calculate the velocity of sound in liquids and solids. Since these bodies have but one voluminal elasticity, an adiabatic expansion or compression is impossible, hence the substitution of the simple values of the elasticity and the density in this expression gives the velocity at once. Determined in this way, the calculated values of the velocity of sound in liquids and solids agree closely with the measurements made by direct experiment.

555. The Echo. A spherical sound wave may be reflected from a distant mountain, wall, or forest, very much as a wave of light is reflected from a mirror. The echo is the repetition of a sound produced by the reflection of this spherical wave by a distant object. If the reflecting wall is so distant that the time taken by the sound wave in advancing and returning is at least equal to the interval required for the perception of two distinct sounds, the echo will return as an independent sound. But if the wall is near at hand, the returning wave will not be widely separated from the direct one, and a more or less serious confusion of sounds is the result.

This is the cause of the disagreeable reverberations which are often heard in large, unfurnished halls. Tapestries and upholstered furniture in a large room, absorb the energy of the reflected sound waves, and so obviate this indistinct echo. The rolling of thunder is simply the echo and the re-echo of the sound wave as this is reflected from cloud to cloud. Whenever sound passes through layers of air of unequal density, reflections occur at the surfaces where the density changes, and the sound wave which is transmitted is weakened by these reflections. It is on this account that sounds can be heard farther by night than by day, because at this time the air is more nearly of uniform density.

CHAPTER XLIX.

MUSIC AND THE MUSICAL SCALE

556. The Characteristics of a Musical Note. When the particles of a medium vibrate, it is found that these particular vibrations may differ from all others in three respects. About a vibration, therefore, there are three features or peculiarities which individualize it or differentiate it from other vibrations, and these may be known as its characteristics. First among these is the *period*, or the time taken for the particle to make one complete vibration. Another is the *amplitude*, or the maximum displacement of the particle from its position of equilibrium. The third is the *law according to which the particle vibrates*, this law being perhaps that of simple harmonic motion, or perhaps one more complex than this.

Inasmuch as a particle vibrating in an elastic medium gives rise to a wave motion in that medium, any peculiarities in the vibration of the particle will cause corresponding individualites to appear in the wave motion which is produced. There are, then, also, *three characteristics of wave motion* depending upon these three characteristics of vibration.

Since the wave length is the distance which the disturbance travels during the time of one vibration, that is, during a single period, it is evident that the length of the wave must depend upon the length of the period. Accordingly, the *wave length* is the characteristic of wave motion which depends upon the *period* of the vibrating particle.

A second characteristic of wave motion is the *amplitude* of the wave, or the maximum displacement which the medium undergoes as the wave passes through it. Since the displacement occurring in the medium is occasioned by the successive displacements of its particles, it follows that the characteristic of *amplitude* in wave motion depends upon the corresponding characteristic of *amplitude* in vibration.

The third peculiarity which differentiates one wave from another is the *form* of the wave. If the particles vibrate with simple harmonic motion, the wave produced will be the smooth, regular curve known as the harmonic curve, or the curve of sines. If, however, the particles vibrate according to some other law, the wave will then have a different *form*. The characteristic of wave motion known as the *form* of the wave depends, therefore, upon *the law according to which the vibration of the particles takes place.*

When the elastic medium in which the disturbance occurs is the air, the vibrations which are produced give rise to the sensation of sound. There are, in general, two heads under which all sounds may be classified, namely, noise and music. When the vibrations of the air occur in a perfectly irregular manner and wholly without periodicity, the ear is thrown into an irregularity of motion corresponding to the vibrations which reach it, and the sensation of *noise* is perceived. A cab rattling over a stony street may be mentioned as an example of pure noise. When, however, the vibrations of the air are regular and periodic, the ear is likewise thrown into regular vibrations, and the sensation of a *musical note* is produced. A *noise,* therefore, is a sound occasioned by *irregular* vibrations, while a *musical note* is the result of vibrations which are *periodic.* In this connection we shall consider musical sounds only.

Having learned that a musical note is produced by regular or periodic vibrations of the air, we must next inquire in what respect the differences which exist in the sound waves in the air give rise to corresponding differences in the musical note. Since there are three characteristics of wave motion, there are also *three characteristics of a musical note,* depending upon the three ways in which waves can differ from one another.

The *wave length* of the sound wave produces that characteristic of a musical note which is known as *pitch*. The pitch of a note is that distinguishing characteristic of a musical note which *determines its position upon the musical scale. Pitch depends upon the length of the wave,* that is, upon the *number of vibrations per second,* and is absolutely independent of the loudness of the note, and is, therefore, also independent of the amplitude of the vibrating particle or of the wave.

The amplitude of the wave affects the excursions which the little air particles have to make as they vibrate back and forth, and these in turn act upon the drum of the ear, causing the latter to vibrate back and forth through wide distances when the amplitude is great, and through small distances when the amplitude is small. The amplitude of the drum of the ear has to deal with the sensation of *loudness* in the sound which is heard, hence the *loudness* of a sound is that characteristic which depends upon the *amplitude of the wave.*

The third feature of a musical note which distinguishes it from any other is called the *quality,* or the *timbre,* of the note. The quality, or the timbre, of a note is that characteristic of a musical note which depends upon the *form of the wave.* One sensation will be produced upon the ear if the form of the wave is that of the simple sine curve, while another will be produced if the wave is of a different form, although both waves may have the same wave length and also the same amplitude. It is the *quality* of a note that distinguishes one person's voice from another, or that causes a note played upon one instrument to sound entirely different from the same note when played upon another instrument. The further study of the matter of quality must be reserved for a later time.

The foregoing conclusions as to the characteristics of vibrations, of waves, and of musical notes, and their dependence one upon the other, may be tabulated in the following form:

Characteristics of Vibration	Characteristics of Wave Motion	Characteristics of a Musical Note
Period	Wave Length	Pitch
Amplitude	Amplitude	Loudness
Law of Motion	Form of the Wave	Quality or timbre

557. The Loudness of Sound. As has been pointed out, the *loudness* of a sound depends upon the *amplitude* of the little air particle which strikes the ear, and therefore upon the amplitude of the movement of the ear drum. If the pitch of a note remains constant, the little air particle, which will have to make one complete vibration in the period, will have farther to move, and therefore will move with higher velocity, if the amplitude is great than if the amplitude is small. But

$$\text{Kinetic energy} = \frac{1}{2} m v^2,$$

where m is the mass of the particle and v is the velocity with which it is moving. Accordingly, the energy of a particle which moves through a great amplitude, that is, with a high velocity, is greater than that of a particle moving through a small amplitude, or with a low velocity. The loudness of a sound depends, therefore, upon the energy of the air particles which strike the ear.

Moreover, the energy of the vibrating particle depends upon the *square* of its velocity, so that if one particle is moving with double the velocity of another, that is, if it moves through twice the amplitude of a second particle, the first has four times the energy of the second and produces a sound four times as loud. Accordingly, if the amplitude of vibration is doubled, the loudness of the resultant sound is increased four-fold. *The loudness of a sound is therefore proportional to the square of the amplitude of the vibrating particle.*

558. Pitch. The *pitch* of a musical note is that characteristic of the note which *determines its position upon the musical scale*. As has been already shown, this depends upon the *length of the wave*, or upon the *number of vibrations per second*. If a succession of regular vibrations or periodic disturbances is made, the ear detects these sensations as individual sounds until they exceed in frequency about 30 per second, after which they blend together forming a musical note, the pitch of which depends upon the number of vibrations per second.

This is nicely illustrated by the device known as Savart's wheel, which consists of a revolving toothed wheel, against which is held a card. As the wheel is turned slowly, the card snaps from tooth to tooth, producing a regular succession of sounds. As long as these vibrations of the card do not exceed about 30 per second, they may be recognized as individual blows against the teeth of the wheel; but when the velocity of rotation increases so that more than 30 teeth per second pass under the card, the ear is no longer able to recognize these blows as distinct sounds, but detects, instead, a note of low pitch, into which these vibrations all blend. As the velocity of rotation is still further increased, the vibrations of the card become more and more frequent, and the note thus produced rises in pitch until it becomes very shrill. Indeed if the number of teeth on the wheel and the velocity of rotation are known, the number of vibrations per second which the card makes, and therefore the number of vibrations per second of the resulting note, may be calculated.

Owing to many irregular vibrations of the card, the tone produced by the employment of Savart's wheel is so far from "pure" that Seebeck was led to devise the instrument known as the siren for determining the number of vibrations per second of any musical note. Seebeck's siren consists of a disc, Fig. 366, with holes at regular intervals around its circumference, and so arranged that it may be rapidly rotated. A jet of air which is held near the circumference opposite the row of holes becomes broken up into a series of regular puffs which pass through the openings. The air, therefore, on the farther side of the disc is thrown into a series of regular vibrations, the number per second depending upon the number of puffs per second, and this latter number may be controlled by regulating the speed of the wheel. These puffs soon blend together into a musical note, the pitch of which may be raised or lowered at will, and the number of vibrations per second is at once determined by knowing the number of holes in the disc and the velocity of rotation. When, therefore, it is desired to learn the number of vibrations per second of a note of any

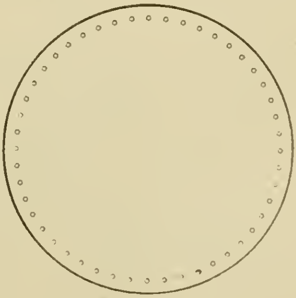

FIG. 366. Seebeck's siren disc.

given pitch, the pitch of the siren is raised until it agrees with that of the note to be determined, and the number of vibrations per second made by the siren is thus ascertained. Since the pitch of the two notes is the same, the number of vibrations per second must also be the same, and thus the number of vibrations of a note of any pitch may be determined.

559. The Siren. Cagniard de Latour improved the siren of Seebeck by so arranging the mechanism that the pressure of the air which passes through the disc and forms the periodic puffs also turns the disc. His apparatus consists in the main of a circular box *A*, Fig. 367, into which air is forced, the top of the box *B* being pierced by a number of oblique holes. Above this, the horizontal plate *C* is so mounted as to turn very freely upon a pivot. This plate *C* is likewise pierced with inclined holes, the holes, however, sloping in a direction opposite to those in the top of the box *B*. When air from a bellows is pumped into the air chest *A*, it is forced out through the inclined holes in *B*, and strikes against the walls of the holes in the plate *C*, thus actually blowing this plate around, and making of it, indeed, a little wind-mill. While the two holes are opposite each other a puff of air passes through, but this is soon intercepted by the turning of the plate. The inertia of the latter, however, carries it on until another hole comes opposite the lower hole, when another puff of air comes through and the plate is again blown around, and so the process continues. The velocity with which the air is driven through these holes, and consequently the speed at which the revolving plate is made to move, may be adjusted by regulating the pressure

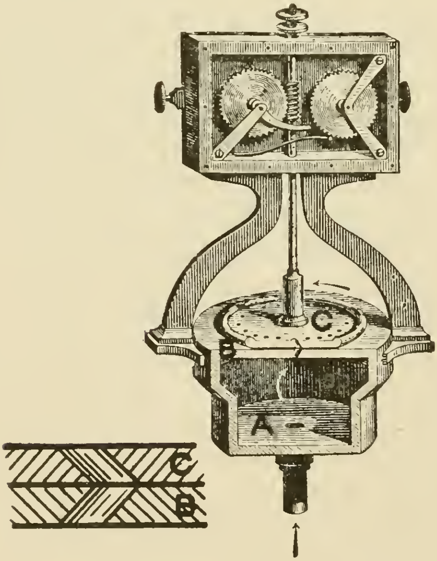

FIG. 367. The siren of Cagniard de Latour.

with which the air is forced into the box *A*, and thus the number of puffs per second which pass through the upper plate is wholly under control.

As the instrument is started, the ear detects at first only a series of separate puffs, but these come faster and faster until after the limit of 30 per second is reached they blend into a note of very low pitch. As the velocity of rotation increases, the number of puffs per second likewise increases, and the note rises steadily in pitch. A simple device attached to the axis of the revolving plate serves to count the number of rotations; and by knowing the number of holes in the disc, the number of puffs per second, and therefore also the number of vibrations per second, may be accurately determined. The pitch of the siren is then raised until it coincides with that of the note the number of vibrations of which is to be determined, and the counting device set in operation. From the registered number of rotations and from the number of holes in the rotating disc, the number of vibrations per second of the note emitted by the siren may be ascertained; and as the pitch of the siren coincides with that of the note under examination, the number of vibrations per second of the latter note may thus be determined. This is the usual method of determining the number of vibrations per second made by a tuning fork or other sounding body.

560. The Range of the Human Ear. It has been shown that the ear is able to recognize individual pulses so long as these occur less frequently than about 30 per second. At a frequency higher than this, these pulses blend together into a sensation free from flutter, though the true conception of pitch is not attained until this number increases to about 40 per second, which corresponds to the lowest note emitted by deep-toned musical instruments. As the frequency increases continuously above this number, the ear recognizes a steadily rising pitch until the highest musical notes have been reached, after which the shrillness of the sound produced becomes painfully unpleasant.

The extreme limit for the perception of sound may be placed, perhaps at about 32,000 vibrations per second, though for all but the best ears a number smaller than this by half is barely perceptible. The number of vibrations per second occurring in the orchestra ranges from 40 for the deepest base to 4752 for the piccolo flute, so that the number available for ordinary music may be placed roughly between 40 and 4000 per second. The range of the musical ear, extending thus through about 7 octaves, is therefore much wider than that of the eye, which is sensitive only to ether vibrations which are comprised within the limits of a single octave.

561. Motion of the Sounding Body Towards or Away From the Observer—Doppler's Principle. When the sounding body and the observer move rapidly with reference to each other in the direction of the line joining them, the pitch of the note is changed, being raised when the sounding body and the observer approach each other, and lowered when the two recede. The principle upon which this effect depends may be made much clearer by considering an illustration given by Professor Tait.

If we think of a stream of water flowing at the rate of three feet per second, and imagine a person standing upon the bank throwing in corks at the rate of one per second, the corks will float down the stream and pass an observer stationed below at a distance of three feet apart. If the person now walks down stream at the rate of one foot per second and continues to throw in corks at the rate of one per second, the first cork will have moved three feet down stream when the second is thrown in; but in the mean time the person himself will have moved down stream one foot before he throws in the second, hence the second cork will be but two feet from the first as they float down stream, and an observer stationed below will see the corks float past him only two feet apart. If now the person turns and walks up stream at the rate of one foot per second, still throwing in corks as before, the first cork will have floated down stream three feet before the second is thrown in; but meanwhile the person has walked up stream one foot, therefore the second cork is distant four feet from the first, and the observer stationed below perceives the corks floating past him four feet apart. The person throws in corks with perfect regularity, whether he stands still, or walks down stream, or up stream. The distance apart of the corks, however, as they float down stream, is very different in the three cases. If the person walks down stream throwing in corks at the rate of one per second, the effect upon an observer stationed below is the same as if the corks were thrown in at a more rapid rate, and if the person walks up stream, the effect upon the observer below is as if the corks were thrown in at a slower rate.

If we apply this illustration to the case of the relative motion of the sounding body and the observer, we may think of the corks which are thrown upon the stream as the vibrations emitted by a sounding body. If the person who throws in the corks stands still, a certain number of corks per second will pass an observer stationed below; and in the same way if the sounding body remains stationary, a certain number of vibrations per second will strike the ear of an observer, producing the impression of a sound of a certain pitch. If, now, the person walks down stream still throwing in corks, the corks will pass the observer closer together, that is, a greater number per second will pass him. So also, if a sounding body moves toward an observer, more vibrations will strike his ear in a second of time, that is, the interval from vibration to vibration will be shorter, and he wll have the impression of a sound of higher pitch. But if the person who throws in the corks walks up stream, the corks are farther apart as they pass the observer, that is, a smaller number will float past him in a second of time, and he has the impression that they are thrown in more slowly. In the same way, if the sounding body recedes from the ear, the interval from vibration to vibration is longer, that is, there are fewer vibrations per second than are produced when the sounding body is stationary, hence the observer perceives a sound of lower pitch.

A motion of the sounding body toward the observer, therefore, raises the pitch of the note, while motion away from the observer lowers the pitch. This may, be perceived to excellent advantage when a locomotive, with whistle or bell sounding, approaches or recedes from the train in which the observer may be seated. At the instant when the locomotive passes and now moves away, the pitch of the whistle or the bell becomes perceptibly flat, showing that fewer vibrations per second reach the ear, and therefore that the sound is of lower pitch. A slightly smaller change of pitch is produced when the observer instead of the sounding body moves in the direction of the straight line joining the two.

This effect of the change in the pitch of a note, due to the relative motion of the source of sound and the observer, is known as the Döppler principle. Another beautiful illustration of this principle will be pointed out under the subject of light.

562. The Musical Scale. From very early times and by all civilizations it has been recognized that certain successions of musical sounds produce pleasing sensations upon the ear. Moreover, even in primitive ages it was perceived that sentiments and emotions could be expressed by musical notes; and with the ancient Greeks especially, recitations were accompanied by a kind of declamatory music, in which notes resembling the modulations of the human voice were used to assist in the delineation of feeling. In the sixth century B. C., Pythagoras determined the succession of eight notes which produce a pleasurable sensation when sounded one after the other, and which are known as the musical scale. Not only this, but, by experimenting with stretched strings of different length, he determined the relations which must exist between these notes to produce this pleasing effect. So satisfactory and sufficient did this Pythagorean scale prove to be, that it remained in general use until the sixteenth century, and even then underwent only a gradual development into the two scales, the major and the minor, which are in use at the present day.

As the major scale now exists, it consists of a succession of eight notes, designated by the letters of the alphabet, beginning with the letter C, as follows:

C D E F G A B C′

In music the first note of the scale gives character or tone to the entire scale and to the music played therein, and is known as the keynote, or tonic. We thus have the scales of C, D, E, etc. The pleasurable sensation which results when the eight notes of the scale are sounded successively, or when certain combinations are played together, depends only upon the relative vibration-frequencies of these notes and not upon their absolute number, as may be observed when a certain group is played in one octave, and when the same group is played in the octave above. In both instances, the ear recognizes the same pleasing sensation, although the actual number of vibrations per second made by any note is twice as great in the latter case as in the former.

The ratio of the vibration-frequencies of any two notes is called the *interval* between these notes. The interval is always expressed as a fraction, of which the vibration-frequency of the higher note is the numerator and that of the lower note is the denominator. It has been found that for every vibration per second made by the first note of the scale, the relative vibration-frequencies of the succeeding notes may be indicated by the following fractions:

C	D	E	F	G	A	B	C′
1	$\frac{9}{8}$	$\frac{5}{4}$	$\frac{4}{3}$	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{15}{8}$	2

These fractions indicate, therefore, the musical interval which exists between the first note of the scale and the succeeding notes. That is, for every vibration which C makes, D makes 9/8 as many, or, while C is making 8 vibrations, D is making 9. In the same way, for every vibration which C makes, E makes 5/4 as many, or, while C is making 4 vibrations E is making 5, and so on. Whatever, therefore, may be the actual number of vibrations per second of a particular note, the eighth note above this will make twice as many vibrations, and is known as the *octave.*

In a similar manner we may determine the interval between each note and the one succeeding. The interval between C and D is 9/8, as we have seen. In the same way the interval between D and E is:

$$\frac{\frac{5}{4}}{\frac{9}{8}} = \frac{10}{9}$$

and so on. The determination of the interval between each note and the note following gives the series of fractions indicated below:

$$\begin{array}{ccccccccccccccc}
C & & D & & E & & F & & G & & A & & B & & C'' \\
1 & & \frac{9}{8} & & \frac{5}{4} & & \frac{4}{3} & & \frac{3}{2} & & \frac{5}{3} & & \frac{15}{8} & & 2 \\
 & \frac{9}{8} & & \frac{10}{9} & & \frac{16}{15} & & \frac{9}{8} & & \frac{10}{9} & & \frac{9}{8} & & \frac{16}{15} &
\end{array}$$

An examination of these fractions shows that there are only three different intervals between the successive notes of the octave, and that these intervals are arranged in a perfectly definite order. The interval between C and D is not the same as that between D and E, while both of these are very different from that between E and F. The interval between C and D is known as a major whole tone, that between D and E a minor whole tone, while the interval between E and F is called a major semitone. Another interval of great importance occurs in music, although it is not found in the simple scale, and that is the interval $\frac{25}{24}$. This is known as the minor semitone.

The succession of eight notes which we recognize as the musical scale must be so related to one another that the seven intervals between them must be in the order of the last named series of fractions. If the first note of the scale were always C, these eight notes would then be already in the order necessary to give the pleasurable sensation, and no further notes would be needed. When, however, some other note, as for instance D, is made the tonic, then these intervals do not succeed one another in the proper order, and other notes, known as the sharps and flats, must be inserted in order that the same series of intervals may follow and the same harmonious effect be produced. This insertion of additional notes between the original notes of the scale consists in raising certain notes by a minor semitone. Raising a note to its sharp is effected therefore by multiplying the vibration frequency of that note by $\frac{25}{24}$, and lowering a note to its flat consists in multiplying the vibration frequency by $\frac{24}{25}$. Accordingly, in order to use each of the seven notes of the scale as a tonic, it is necessary to provide each note with a sharp and with a flat, thus making twenty-one notes in all. Moreover, since it is possible to use each sharp and also each flat as a tonic, each sharp must itself be provided with a sharp, and each flat likewise provided with a flat; so that in order to maintain the strict relationship between the relative vibration frequencies when each note in turn is used as the tonic, it is necessary to have thirty-five notes in the octave. Indeed, Helmholtz constructed an harmonica having this theoretical number of keys, and it was found that the same perfect succession of sounds was produced whatever note of the scale was used as a tonic.

On account, however, of the unwieldiness and the difficulty of playing so elaborate an instrument, it was found necessary very early in the history of music to limit the number of keys in the octave, and twelve in all seemed the most feasible number. Accordingly, between every two notes a major whole tone or a minor whole tone apart, there was inserted a single extra (black) key instead of the four which properly belong there, and the pitch of this note was so adjusted that, while it was not perfectly correct for any one scale, it would do for all without making very serious discord. The eight principal notes of the scale were then likewise adjusted until they would produce the least discord with the additional half tones which had been inserted. This process thus of changing the vibration-frequencies of the notes of the scale so as to obtain the least discordant result when the notes of the keyboard were limited to twelve was called *tempering* the scale. Naturally this method of varying and adjusting the notes gave rise to great confusion, and previous to the time of Bach a number of different temperaments were in existence.

As the best means of tempering the scale, Bach proposed to divide the octave into *twelve equal intervals,* that is, so that the interval between any two adjacent notes would be the same as that between any other two adjacent notes. This plan of tempering the scale which was adopted at that time is still in use, and is known as *Bach's even temperament.* Accordingly, on an instrument such as the piano which has a fixed key-board, the octaves are the only notes which are in perfect accord, since all the others have had their vibration frequencies changed a little in order to divide the octave into twelve equal intervals. Music, therefore, which is played upon an instrument with fixed keys is far from perfect; and we have simply trained our ears not to mind the discord which is always present, even when the instrument is in so-called perfect tune. Upon an instrument, however, like the violin, where the number of keys is not limited, but where the pitch of the note depends upon the length of the string, the skillful player, by properly adjusting this length, can give the various tones their theoretical number of vibrations, and thus can produce music far more perfect and of a higher order. It follows, therefore, that if the violin player were to play correctly, he would be out of tune with the accompanying piano.

From the diagram herewith given, Fig. 368, a comparison may be obtained between the perfect scale and the modern scale of even temperament. In many instances it is evident that the discrepancy between the two is wide, and that modern music must lack greatly in sweetness and character. A return to the perfect scale is therefore earnestly to be desired; and although a certain brilliancy of execution might be sacrificed because of a more complicated key-board, the greater beauty and harmony which would then result would more than compensate for this small loss.

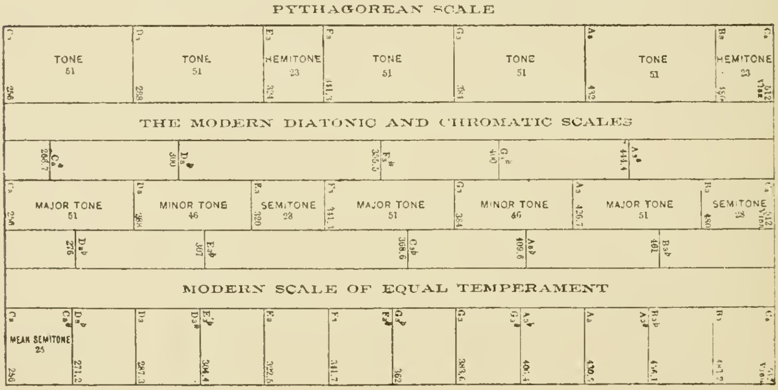

FIG. 368. Comparison between the Pythagorean scale, the perfect major scale with its sharps and flats, and the modern scale of even temperament.

The modern minor scale is formed likewise of eight notes, and consists of the following relative vibration-frequencies and intervals:

$$\begin{array}{ccccccccccccccc} 1 & & \frac{9}{8} & & \frac{6}{5} & & \frac{4}{3} & & \frac{3}{2} & & \frac{8}{5} & & \frac{9}{5} & & 2 \\ & \frac{9}{8} & & \frac{16}{15} & & \frac{10}{9} & & \frac{9}{8} & & \frac{16}{15} & & \frac{9}{8} & & \frac{10}{9} & \end{array}$$

It is thus seen that the minor scale is composed of the same intervals as the major scale, but differently arranged. Strangely enough, another form of the minor scale exists, of which the first half consists of the minor scale given above, while the second half is formed from the major scale, as follows:

$$\begin{array}{ccccccccccccccc} 1 & & \frac{9}{8} & & \frac{6}{5} & & \frac{4}{3} & & \frac{3}{2} & & \frac{5}{3} & & \frac{15}{8} & & 2 \\ & \frac{9}{8} & & \frac{16}{15} & & \frac{10}{9} & & \frac{9}{8} & & \frac{10}{9} & & \frac{9}{8} & & \frac{16}{15} & \end{array}$$

The latter arrangement of notes is preferred for the ascending minor scale, while the former is used for the descending scale.

The difference between the major and the minor scales in the impression produced upon the ear seems the more remarkable when we remember that this distinction is brought about only by a rearrangement of the successive intervals. The major scale as well as the music written therein is characterized by a certain freedom, clearness, tranquility, happiness, while music in the minor scale depicts suffering, remorse, unhappiness, and especially doubt and indecision.

Whenever two or more notes are sounded together, the sensation is harmonious or discordant according to the relative vibration-frequencies of the notes. It is found that if the relative vibration-frequencies may be expressed by small numbers, the resulting sound is pleasing; but if these can be expressed only by large numbers, the resulting sound is thoroughly disagreeable. The farther the relative vibration-frequencies depart from unity, that is, the larger the numbers which express the relative number of vibrations, the more acute is the discord.

Thus, a note and its octave, the relative number of vibrations of which are 1 and 2, give a perfect accord, since the numbers expressing the relative vibration-frequencies are small. So, also, when C and G are sounded together, we have a very smooth, harmonious resultant sound, since the numbers expressing their relative vibration-frequencies, namely 2 and 3, are likewise small. In the same way, C and E give a pleasurable sensation when sounded together, although, the trained ear detects a slight roughness not noticed in either of the other combinations. If, however, E and G are sounded together, we have here two notes which have vibration-frequencies represented by 5 and 6, and the resulting tone has a roughness which differs but little from discord. The case is still more extreme when C and D or when B and C are sounded together, the sound then being exceedingly disagreeable.

Four notes, C, E, G, C′, having the relative vibration-frequencies 4, 5, 6, 8, gives a particularly agreeable sound which is known as the major chord. Four other notes, A, C, E, A′, having the relative vibration-frequencies 10, 12, 15, 20, likewise form a pleasurable combination known as the minor chord. The cause of the greater uncertainty and unrestfulness depicted by the latter chord is found in the physical fact that the relative vibration-frequencies are expressed by much larger numbers. In writing music, the composer unconsciously expresses his thoughts in combinations of notes which have among themselves small or large relative vibration-frequencies, according as he wishes to express sentiment of a harmonious or discordant nature. Thus happiness and contentment are expressed by music written in fifths (the interval from C to G) or in thirds (the interval from C to E), while pain, unhappiness, discontent, suffering, etc., are expressed by music written in sevenths (the interval from C to B). The reason as to why sensations are harmonious or discordant according as the relative vibration-frequencies are small or large will be better understood later, after we have examined the resulting form of the wave actuated by two vibrations.

In all that has been said about the musical scale, no reference has yet been made to the actual number of vibrations which produce the various notes. At the time of Bach and Händel, the number of vibrations per second made by the note A was 422.5, and at the beginning of the eighteenth century it was as low as 405. It was found, however, that a certain brilliancy is added to music by raising the standard of pitch, and in 1857 in the

European opera it reached the number 448. Because of the difficulty in singing at so high a pitch, this number has since been reduced until it is now 435. A standard A-tuning fork making 435 vibrations per second has been deposited in the conservatory at Paris, and this has been universally adopted as the standard of pitch. In physics, however, the number of vibrations per second of middle C on the scale has been retained at 256, which is a very convenient number inasmuch as it is the eighth power of 2, although this note is of lower pitch than the corresponding note of the scale in which music is now universally played.

CHAPTER L.

RESONANCE.

563. Sympathetic Vibrations. If we have a heavy cannon ball suspended by a stout rope, and attempt to move it out of its position of equilibrium by striking it, we shall find that even quite a severe blow will have but little effect upon it. If we now attach to it a slender thread—a thread so delicate that it would not support the one-hundredth part of the weight of the ball—and exert a slight pull upon it, the effect produced will be almost imperceptible. If, however, we pull upon the thread at regular intervals, the time between pulls being exactly that required for the cannon ball to make one complete vibration, we shall find that the first pull causes it to move forward and back through a distance so small that it can hardly be detected. The ball, however, does execute a vibration of very small amplitude; and if the second pull comes just as the ball swings forward the second time, the effect of this second pull is added to that of the first, and the amplitude of the vibration is increased. So, also, if other pulls come at exactly equal intervals, each pull coming just as the ball is moving forward, the cumulative effect of these all but imperceptible impulses will be to produce a vibration through a very considerable amplitude.

If, instead of pulling the ball by a bit of thread, a puff of air is directed against it, the effect of a single puff is very slight indeed in the motion which it communicates, but it nevertheless causes the ball to vibrate through a small amplitude. If now a second puff comes just after the ball has completed the minute vibration caused by the first, the effect of the second is added to that of the first, and the ball vibrates with greater amplitude. In the same way, if a whole series of puffs is directed against the ball, each one coming just as the ball is about to begin a new vibration, the effects of the puffs are additive, and the ball is soon thrown into a vibration of comparatively large amplitude. Everything in this experiment depends upon the fact that the pulls upon the thread or the puffs of air against the ball must be of *exactly the same period as the vibration of the ball itself,* so that each impulse, although slight, may come in the proper phase to re-enforce the vibrations which have already been imparted.

When one vibrating body thus causes another of *the same period as itself* to vibrate, the vibrations of the latter are said to be *sympathetic.* A beautiful illustration of sympathetic vibrations is shown by a number of pendulums suspended from a frame. If one pendulum is set in vibration, it will impart a slight movement to the supporting frame, this small motion of the frame coinciding in period with the vibration of the pendulum. All the other pendulums will then receive a slight impulse, tending to make them swing back and forth. If their periods, however, are different from that of the first, these small impulses which are communicated through the frame will not come at the right time to produce an additive effect, hence no vibration will be imparted to the others. But if there is one among this number which has the same period as the first, a second impulse which is transmitted through the frame will increase the slight vibration produced by the first. Other impulses will then follow in exactly the right phase to make the motion each time a little greater, hence this pendulum will in time acquire a very considerable motion. One vibrating pendulum, therefore. will set in vibration another which has the same period as itself, if the two are so mounted that the impulse from one may be communicated to the second. Since the motion of the second represents a certain amount of energy, and since this motion is derived from that of the first, the second pendulum is thrown into vibration at the expense of the energy of the first, and the first is brought quickly to rest. It is now possible for the same process to be repeated in the reverse order, the second pendulum imparting motion to the first, and being itself brought to rest.

What has been said of sympathetic vibrations, is also true when the bodies are much smaller than the massive ones here considered, and when the impulses are far more delicate than the mechanical pulls or puffs or movements here imagined. If one tuning fork is caused to vibrate and is placed near another of exactly the same pitch as itself, the periodic condensations of the air caused by the first fork strike the second with perfect regularity, and produce upon it the same effect as the periodic puffs of air produced upon the cannon ball. The first condensation of the air causes a slight back and forth excursion of the prongs of the second fork, and after this completes its first very small vibration, another condensation comes just as the prongs are about to begin their second vibration. The effect of the second disturbance is added to that of the first, and the fork vibrates with a greater amplitude. This process continues until the amplitude of the second fork is sufficiently great to produce a sound loud enough to be heard. Indeed, if the two forks are widely separated, and one is made to vibrate by bowing or striking, the other will be heard almost immediately. If the second is bowed, the first will respond in like manner.

So also if the two strings of a sonometer are tuned to unison and one is made to vibrate, the other is thrown at once into vibration. If the first is then stopped, the second will be heard emitting a note of the same pitch as the first. Indeed, if the first is made to vibrate in two segments so as to give the first harmonic of the fundamental, the second will be thrown into similar vibrations, as may be shown by the fact that little paper riders placed upon this string will be thrown off except the one in the middle, where there is a node. If the first string subdivides into three segments, the second may be shown by the paper riders likewise to be vibrating in three segments and emitting its second harmonic, and so on. In the same way the phenomenon of sympathetic vibrations may be beautifully shown by having two Chladni plates exactly alike, and setting one in vibration by bowing. The sand will gather along the nodal lines, showing that vibrations of a certain character exists upon the plate. If sand is now scattered upon the second plate, and this is held near the first, the second will be thrown into vibrations similar to the first, Fig. 369, as will be shown by the sand pattern upon the second plate appearing identical with that on the first.

A stretched string, or a vibrating reed, or a column of air, since its own period of vibration is fixed, can be thrown into sympathetic vibrations only when a note of exactly the same pitch as that which it itself emits is sounded near it. A stretched membrane, however, responds to any sound and vibrates sympathetically with it. This is undoubtedly due to the fact that a membrane readily divides into smaller vibrating segments, so that when almost any sound is made before it, it so subdivides that it is thrown into vibrations which coincide in period with those made near it. In other words, it is only to a very slight extent that a stretched membrane has a period of vibration peculiar to itself. A membrane, therefore, is very ready to break up into a series of vibrating segments separated by nodal lines; and the particular combination of vibrating segments is different whenever different sounds are made before it. In this way the drum of the ear vibrates in unison with any sound that is made before it, thereby enabling us to hear sounds which differ greatly in pitch and quality. Furthermore, it is upon this principle that the action of the telephone depends, for here, likewise, we have a diaphragm which vibrates in unison with almost any sound. In a few instances, however, the telephone fails to vibrate sympathetically with sounds made before it, and therefore, as is well known, there are certain sounds which this instrument does not transmit as readily as others.

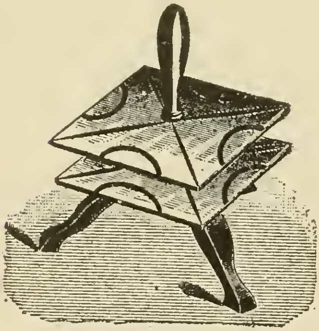

FIG. 369. Sympathetic vibrations shown by Chladni plates.

564. Forced Vibrations. The phenomenon of sympathetic vibrations occurs, as we have seen, only when the natural or free period of the second body is exactly the same as that of the first. Each impulse received from the first then coincides in phase with the vibrations which have already been imparted to the second, so that the latter is thrown into vibrations of its own free period.

The situation is different, however, when the first vibrating body imposes its own motion upon the second, although the free periods of the two may differ widely. In this case the second is *forced* to vibrate with the period of the first, giving rise to the phenomenon known as *forced vibrations.* This may be illustrated by suspending a small pendulum from a long, massive one, Fig. 370, and setting both in vibration. The smaller pendulum, though tending to vibrate with a much shorter period, is nevertheless carried along by the larger one, and after an interval of irregular motion, the two reach a condition of equilibrium in which the smaller is compelled by the vibrations which are thus impressed upon it to execute a forced vibration which is very different from its own free period.

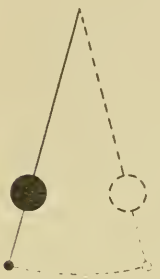

Fig. 370. The smaller pendulum vibrates with forced vibrations.

In this way a tuning fork causes forced vibrations to occur in the resonating box upon which it is mounted. Similarly, a vibrating tuning fork, held against a door, throws the latter into a forced vibration of the same period as itself, whereby the volume of the sound emitted by the fork is greatly increased. The same effect is produced when the fork is held against the end of a long rod which itself rests against the door, the entire system of rod and door then being compelled to vibrate with forced vibrations in the same period as the fork. It is upon the principle of forced vibrations that the sounding board of a musical instrument is made to vibrate in unison with the strings which rest upon it. By its own vibration and that which it imparts to the adjacent air, it greatly reenforces the sound from the strings, which would otherwise be feeble and ineffectual.

565. The Reenforcement of Sound—Resonance. When one vibrating body throws another into sympathetic vibrations, the latter body by its motion sets in vibration a larger mass of air, and the intensity of the sound which is thus produced is greatly increased. This reenforcement of a sound by the synchronous vibration of another body is known as *resonance*. A column of air of a particular length may be thrown into vibration by a sound of a definite pitch, so that this column vibrates in unison with the sound and reenforces it. Such a column of air, therefore, serves as an excellent resonator for a sound of one particular pitch.

This may be nicely illustrated by removing a vibrating tuning fork from its resonating box and holding it over the mouth of a tall, narrow jar, at the same time pouring in water, Fig. 371. The fork by itself is almost inaudible; but when the column of air in the tube above the water has reached a perfectly definite length, the vibrations of the fork are strongly reenforced, and the former feeble sound becomes one of great volume. If more water is added, thus shortening the column of air by ever so little, the intensity of the sound diminishes at once, showing that the perfect resonance, and therefore, also, the synchronous vibration of the column of air with the tuning fork, can occur only when the air column is of a perfectly definite length. What this length is for a note of a certain pitch may be determined by analyzing the disturbances which take place in the air column during one complete vibration of the fork. For the sake of simplicity, we shall consider the vibration of but a single prong, and shall further assume that the column of air has been adjusted to such a length that the greatest reenforcement of the sound occurs.

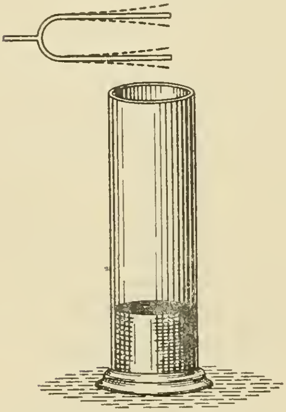

Fig. 371. The resonance of a column of air.

566. The Fundamental Vibration of a Column of Air Closed at One End. Let us assume that o, Fig. 372, is the position of equilibrium of the prong, and that the fork vibrates through the amplitude *oa* or *ob,* that is, in one complete vibration the prong moves from the position *a* down to the position *b* and back again to *a*. If then the prong is at *a* and just about to move, the motion downward condenses the air immediately beneath, this in turn condenses the air farther down, and so a wave of condensation moves down the tube. At the bottom this wave of condensation is reflected with change of sign (p. 665), and now moves as a condensation up the tube, reaching the mouth just as prong has reached its lowest position *b*. That is, during the time in which the prong has been moving from *a* to *b,* the wave of condensation has travelled down the tube and back again to the mouth.

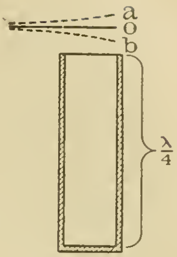

FIG. 372. Fundamental vibration of column of air closed at one end.

Since the disturbance can move more freely in the outside air than in the air enclosed by the rigid walls of the tube, a little air particle in this wave of condensation, which has now returned to the mouth of the tube, moves farther than any other particle. The larger movement of this little air particle causes a rarefaction to be formed behind it, and this likewise is propagated downward. In other words, this condensation, which has travelled down the tube and back again to the mouth, meets here the outside air, and in this it can move with greater freedom than in the air enclosed by the walls of the tube in which it has been travelling. Reflection, therefore, *without change of sign* (p. 666) but *with the loss of half a wave length* occurs at this point; and the condensation which has been travelling up the tube is reflected at the mouth as a rarefaction which now starts down. Moreover, this has all been accomplished while the prong has been moving from *a* to *b*, so that the prong is now about to move upward.

As the prong moves upward, a rarefaction is left behind it, the particles of air immediately beneath rush into this rarefaction, leaving a rarefaction behind them, and in this way a wave of rarefaction is propagated down the tube. Since this rarefaction produced by the upward movement of the prong coincides with the one produced by the reflection of the condensation at the mouth of the tube, a rarefaction of increased amplitude travels down the tube. At the bottom it is reflected with change of sign and comes back as a rarefaction, arriving at the mouth of the tube just as the prong of the fork reaches the position *a*. Accordingly, while the prong has been moving from *b* to *a*, a wave of rarefaction has travelled down the tube and back again to the mouth.

It is thus seen that while the fork has been making one complete vibration, the disturbance has travelled down the tube and back again as a condensation, and down the tube and back again as a rarefaction. In other words, the length of the tube has been travelled four times during the time of one vibration, and therefore, (p. 658) the wave length of the sound is four times the length of the tube. When the rarefaction which has been propagated up the tube reaches the mouth, it is there reflected as a condensation which starts at once down the tube. This agrees exactly in phase with the condensation which the prong sends out as it begins its second vibration downward, and hence these unite in producing a condensation of greater amplitude, which is sent down the tube. In this way the entire column of air within the tube is thrown into vibration, these vibrations agreeing precisely in phase with those of the fork, and thus the intensity of the sound emitted by the fork is greatly increased. If the length of the column of air should differ by ever so little from this particular amount, these vibrations would not occur in perfect unison with those of the fork, and a reenforcement of the sound would not result. When the vibration of a column of air takes place in this way, the column is said to vibrate with its fundamental vibration, and to emit its fundamental note.

567. The First Harmonic of a Column of Air Closed at One End. If the same fork is held over a much longer tube and water is poured in as before, or if water is drawn off from the tube which we have been considering until a much longer air column is left, another point will soon be found when the same reenforcement of the fork again occurs. In this case, the column of air is exactly half a wave length longer than before, as may be seen by following carefully the aerial displacements through two successive vibrations of the prong.

If *A*, Fig. 373, was the position of the surface of the water in the first instance, and *C* is the position of this surface when resonance is again produced, *B* being a point half way between, there is nothing at *A* to reflect the disturbance as before, and the condensation, which starts down the tube when the prong leaves *a*, travels past *A* unimpeded. This condensation, therefore, is reflected at *C* as a condensation and as such travels up the tube. If the tube is of the proper length, the first condensation reaches the point *B* on its journey upward when the prong has made one complete vibration and is now about to begin its second. A second condensation, therefore, starts down the tube when the first is at the point *B* in its upward movement, and the two meet at the point *A*. Accordingly, the air particle at *A* is urged upward by the first condensation and downward by the second, so that this is a stationary point, or a node. The node *A* is thus determined as the place where one condensation coming up the tube meets the next condensation as it travels down the tube.

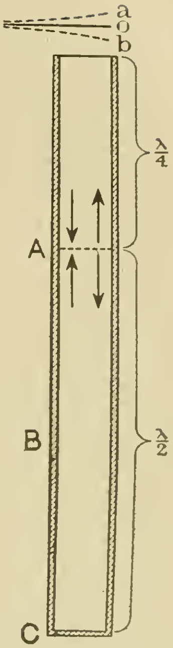

Fig. 373. First harmonic of a column of air closed at one end.

When the first condensation reaches the point *B* as it moves down the tube, the prong of the fork has reached the point *b;* and as it now begins to move upward a wave of rarefaction travels down the tube. This wave of rarefaction is reflected as a rarefaction at the bottom, and so travels as a rarefaction up the tube, reaching the point *B* when the fork has made one complete vibration, and the prong is again in the position *b*, about to move upward for the second time. A second wave of rarefaction, therefore, starts down the tube when the first has reached the point *B* on its upward way; and these two rarefactions, travelling in opposite directions, then meet at the point *A*, again urging this particle in opposite directions. Accordingly, a node in a column of air which is vibrating with stationary waves is alternately the point of meeting of two condensation which are travelling in opposite directions, and of two rarefactions which are travelling in opposite directions. The closed bottom of the tube is necessarily a node, and hence the distance between *A* and *C* is absolutely half a wave length of the sound produced. When a column of air closed at one end is vibrating in the manner here described, having a node at the closed end and also one a third of the length of the pipe from the open end, the column of air is said to be vibrating with its first overtone or harmonic.

This analysis would tend to show that the distance from the open end of the tube down to the point *A* is exactly a quarter wave length long; but practically this distance is shorter than a quarter wave length due to the opening of the tube. The only way in which the length of the wave can be accurately determined by this method is to fill the tube with water to the point *A*, when reenforcement ensues as in the first instance, and to take careful note of the position of this point. The level of the water must then be lowered until reenforcement again takes place, and this determines the position of the node at *C*. The distance between *A* and *C* is then exactly half a wave length.

568. The Vibration of a Column of Air Open at Both Ends. Just as a column of air closed at one end can reenforce the sound emitted by a vibrating reed, so in a similar manner can a tube open at both ends be made likewise to vibrate in resonance with this reed, and thus to reenforce the sound produced. In this instance, however, the length of the column of air must be at least half a wave length long, as may be seen from the following consideration. Suppose, as before, that the prong is at the point *a*, Fig. 374, about to begin its movement downward. While the prong is moving from the position *a* to the position *b*, the wave of condensation travels the whole length of the tube. At the lower end the last air particle can move more freely than any of the others, since the latter are hemmed in by the walls of the tube, and this last particle therefore moves out much farther, leaving a rarefaction behind it. In other words, the wave of condensation, as it reaches the lower end of the tube, is reflected as a rarefaction which travels up the tube. At this instant, however, the prong *b* starts to move upward, thus sending a wave of rarefaction down the tube. This wave of rarefaction now moving downward meets at *A* the wave of rarefaction moving upward, and a node is determined at this point.

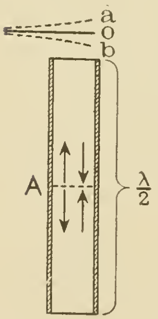

Fig. 374. Fundamental vibration of a column of air open at both ends.

The rarefaction moving up the tube is reflected at the upper end as a condensation just as the prong, which has now reached *a*, sends its second condensation down the tube, and these two accordingly unite in moving down. At this instant, the rarefaction which travelled down the tube is reflected at the lower open end as a condensation which now travels travels up the tube, and two condensations travelling in opposite directions likewise meet at the node *A*. The node *A*, therefore, is the place where alternately two rarefactions meet travelling in opposite directions, and where two condensations meet likewise travelling in opposite directions. The open end of the tube is always an antinode, and the tube itself is half a wave length long. When a column of air open at both ends is vibrating with but a single node at the center, it is said to vibrate with its fundamental vibration.

If now, the tube is lengthened by half a wave length, Fig. 375, a similar analysis of the vibration will show that just as the node at *A* is determined at the distance of one-quarter of a wave length from the upper end, so also another node *B* will be determined at the distance of one-quarter of a wave length from the lower end of the column of air, the distance between these two nodes being half a wave length. The column of air is then said to vibrate with its first overtone or harmonic.

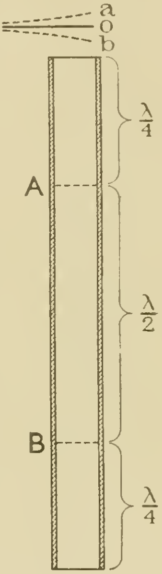

Fig. 375. First harmonic of a column of air open at both ends.

569. The Organ Pipe. The reenforcement of the sound of a certain pitch by a column of air of definite length is the principle of the organ pipe. This instrument consists of a tube, usually of wood or metal, Fig. 376, enclosing a column of air, and having a box, or wind chest, in the lower part into which air can be pumped. This air is then forced through a narrow slit against a thin tongue of the wood or metal forming a part of the side of the tube, and by a process not well understood, this jet of air as it strikes the tongue is thrown into regular vibrations. This vibrating jet or sheet of air is sometimes called the air-reed; and it is the vibrations of this air-reed which are reenforced by the air enclosed within the tube.

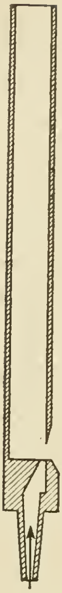

Fig. 376. The organ pipe.

The end of the tube where this vibrating air-reed is situated is necessarily an antinode, and the other end may be closed or open. If closed, we have a closed organ pipe, and a node is determined at that point. If open, we have an open pipe, and the open end must always be an antinode. The number of vibrations per second which the air-reed executes depends upon the pressure of the air which is forced through the narrow slit, and becomes greater as the pressure of the air increases.

570. Fundamental and Harmonic Vibrations in an Organ Pipe. In order to study the formation of nodes and antinodes in the organ pipe, we shall consider first the closed pipe. If the pressure of the air is moderate, and the pipe is of the proper length, the entire column of air enclosed by the pipe is thrown into vibrations having a single node, and this is at the closed end. The pipe is then vibrating with its fundamental vibration, which is the lowest possible note it can emit, and the length of the pipe is 1/4 of the wave length of the sound produced. This is represented in the first diagram, Fig. 377. If the pressure of the air is now increased, the air-reed makes a greater number of vibrations per second, and the note emitted by the pipe suddenly breaks into a clear note of higher pitch. In this instance the pipe emits its first overtone, and the vibrations take place as shown in the second diagram, Fig. 377. The length of the pipe is now 3/4 of the wave length of the sound. A greater pressure of the air causes the air-reed to vibrate still more rapidly. the note then emitted rises in pitch, and the vibrating column is broken into segments as shown in the third diagram, Fig. 377. The length of the pipe is now 5/4 of the wave length of the sound, and so on.

In the case of the open pipe, the fundamental vibration is produced when but a single node is formed; and since this is at the middle of the pipe, the length of the pipe is ½ of the wave length of the sound produced. This is shown in the first diagram, Fig. 378. An increased pressure of the air causes the pipe to vibrate with its first overtone, the vibrations then taking place as shown in the second diagram, Fig. 378. Here the pitch of the note is an octave higher than that produced when the fundamental is sounded, and the length of the pipe is 2/2 wave lengths, or just equal to the wave length of the sound produced. A still greater pressure causes the vibrating column of air to break up into segments as shown in the third diagram, Fig. 378, and emit its second overtone, the length of the tube in this case being 3/2 wave lengths, and so on.

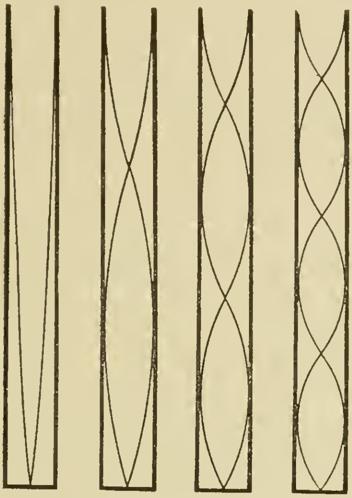

FIG. 377. Fundamental vibration and the first, second, and third harmonic of a column of air closed at one end.

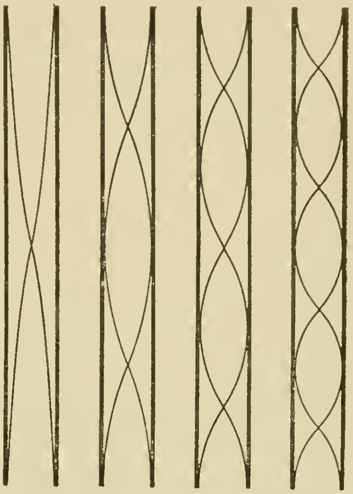

FIG. 378. Fundamental vibration and the first, second, and third harmonic of a column of air open at both ends.

An examination of these figures will show that the length of a closed organ pipe is always 1/4 λ, 3/4 λ, 5/4 λ, 7/4 λ, ..., where λ is the wave length. That is, the length of the *closed* organ pipe is always an *odd* number of quarter wave lengths of the sound produced. In the same way, it will be seen that the open organ pipe is always 2/4 λ, 4/4 λ, 6/4 λ, 8/4 λ, ..., that is, the length of the *open* organ pipe is always an *even* number of quarter wave lengths of the sound. It follows that the vibrations of the closed pipe take place according to the odd numbers 1, 3, 5, 7, ... while those of the open pipe take place according to the even numbers 2, 4, 6, 8, ... Therefore, an open pipe, when vibrating with its fundamental vibration, gives a note an octave higher than a closed pipe of the same length which is also vibrating with its fundamental vibration, since the wave length in the open pipe is only twice the length of the pipe, while that in the closed pipe is four times the length of the pipe. If the hand is laid over the open end of a pipe when this is sounding, the note will drop an octave in pitch.

CHAPTER LI.

INTERFERENCE OF SOUND.

571. The Principle of Interference. Whenever two or more waves exist upon the same medium at the same time, the displacement which the particles actually undergo is the algebraic sum of the displacements which each wave would produce separately. This phenomenon of producing a *resultant displacement* by the simultaneous action of two or more waves is known as *interference*. Interference occurs whenever two or more waves exist upon the same medium at the same time. Accordingly, it is possible for two sound waves to interfere, and the resulting wave is then the algebraic sum of the effects which each would produce by itself. If, therefore, two sounds of the same wave length coincide in phase, the amplitude of the resultant wave is the sum of the amplitudes of the separate waves, and a sound of increased intensity is the result, Fig. 379. But if two sounds of the same wave length give rise to vibrations of opposite phase, the resultant wave has an amplitude equal to the difference between the amplitudes of the two waves, Fig. 380. From this it follows that if these two component sounds, besides being of equal wave length and opposite in phase, are also of equal amplitude, the resultant amplitude is zero, and no sound is heard, Fig. 381. Two sounds, therefore, may interfere so as to produce a much louder sound than either alone, or two sounds under proper conditions may interfere so as to produce silence. It is because of the destructive interference of two sounds, whereby two sounds may combine to produce silence, that we believe in the wave theory of sound.

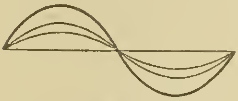
FIG. 379. Resultant of two waves of same wave length, same phase.

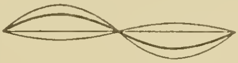
FIG. 380. Resultant of two waves of same wave length, opposite phase, unequal amplitude.

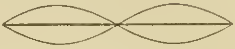
FIG. 381. Resultant of two waves of same wave length, opposite phase, equal amplitude.

A beautiful experiment illustrative of the destructive interference of two sounds consists in holding a tuning fork over a tall jar, and pouring in water until resonance occurs. If the tuning fork is slowly rotated, a point is soon found where the condensation sent down by one prong exactly coincides with the rarefaction sent down by the other, that is, the waves sent out by the two prongs are then in opposite phase, and silence is produced. In this instance two sounds interfere so as to produce silence. If now, a tube of paper is slipped over one of the prongs, thereby intercepting the waves which this prong emits, the waves which are sent out by the other prong are no longer destroyed by interference, and sound is again produced.

572. Beats. If two tuning forks are sounded together, one making 512 vibrations per second and the other 511, the more rapidly vibrating fork will gain one vibration per second over the other. At the beginning of the second they will vibrate together, that is, they will be in the same phase, then after the lapse of half a second they will vibrate in opposite directions, or be in opposite phase, and at the end of the second they will again vibrate in the same direction and be once more in the same phase. At the beginning of the second, when both prongs are moving in the same direction, the two sounds interfere so as to produce a louder sound; half a second later, when the prongs are moving in opposite directions, they interfere destructively so as to produce a diminished sound, and at the end of the second, when the prongs once more vibrate together, they again produce an increased sound. A fluctuation in the intensity of the sound is therefore produced, which gives rise to the phenomenon known as beats. *Beats* are the variations in the intensity of a musical sound, produced when two notes *which do not differ widely from unison* are sounded together. The number of beats per second is the difference in the number of vibrations per second of the two notes. It is to the occurrence of beats that the most serious discord which can occur in music is attributed.

573. Konig's Apparatus for Determining the Wave Length of a Musical Note. By making use of the phenomenon of interference, König devised an apparatus, Fig. 382, whereby the wave length of any musical note may be accurately measured. The sound from the given source, a tuning fork *A*, for instance, after being reenforced by the resonator B, is made to pass through two tubes which divide at *C*. The sound wave likewise divides, one part passing through the branch *CDE*, and the other through the branch *CFG*. At *E* and *G* are manometric capsules (p. 654), so that the vibrations which travel through the upper branch actuate the capsule at *E*, and those which pass through the lower branch act upon the the capsule at G. The gas pipe leading from each of these capsules to the common light *L* causes this flame to be affected by the waves which come through both of the branches of the divided tube. If the distance from *C* to *E* by the path *CDE* is the same

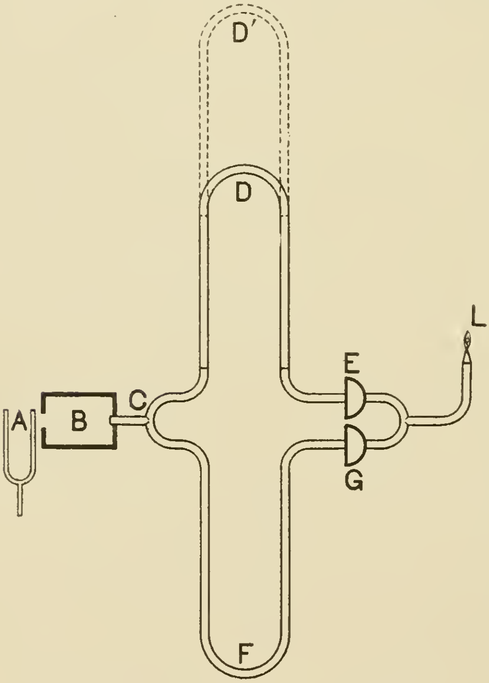

FIG. 382. König's interference apparatus.

as that from C to G by the path CFG, the sound reaches these manometric capsules in precisely the same phase, that is, when a condensation reaches one, a condensation reaches the other, and when a rarefaction arrives at one, a rarefaction arrives at the other. The diaphragms in both capsules then move in the same direction at the same time, and the little flame vibrates with very great amplitude, thus producing a sharply-marked, serrated image, as shown in Fig. 383.

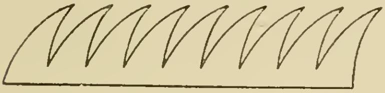

FIG. 383. Manometric flame when the upper and the lower paths are of equal length.

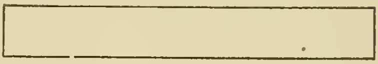

FIG. 384. Manometric flame when the paths differ in length by half a wave length.

The apparatus is so arranged, however, that one branch may be extended, thereby increasing the distance which the sound wave travels in that branch. If, now, this branch is lengthened by drawing out the tube, so that the path $CD'E$ is just half a wave length longer than the path CFG, the waves which reach E and G will be exactly opposite in phase, that is, when a condensation arrives at E, a rarefaction arrives at G. Accordingly, the diaphragms of these capsules vibrate in opposite directions at the same time, and the little flame, which is actuated by both capsules, remains undisturbed. When this occurs, the image of the little flame as seen in the revolving mirror is not serrated, but is simply drawn out into a uniform band of light, as in Fig. 384. The linear distance from D to D' is then exactly one-quarter of a wave length, since the length of this path has been increased by half a wave length.

When the wave length λ of a musical note and also the number of vibrations per second n of this note are known, the velocity of sound v may be calculated at once from the fundamental equation (p. 679).

$$v = n\lambda.$$

574. Composition of Wave Motion. We have seen that when two motions coexist upon the same medium, the actual displacement of a given particle is the resultant of the displacements which each wave would produce separately. When, therefore, the waves are of equal length and occur in the same phase, the resultant wave is likewise of the same length, but of an amplitude equal to the sum of the amplitudes of the component waves. This, accordingly, produces a note of the same pitch, but louder than either of the components. When, however, two waves of equal length occur in opposite phase, the resultant wave is again of the same length, but of an amplitude which is the difference between the amplitudes of the two components. This produces a note of the same pitch as that of either wave alone, but fainter. Indeed, if two waves of equal lengths and equal amplitudes occur in opposite phase, the resultant amplitude is zero, and the two sounds, to which these waves give rise, interfere so as to produce silence.

In all cases, whatever the wave length or the amplitude of the components may be, the *resultant wave* is the one formed by the *resultant displacements* produced by each of the component waves separately. Thus, if we imagine the simple wave *ACE*, Fig. 385, to exist upon a medium, and also another simple wave *ABCDE*, Fig. 386, of half the length of the first to exist upon the same time, the resultant medium at the same displacement which the particles undergo is everywhere the algebraic sum of the displacements which each wave would produce separately. The resultant wave, which is simply the algebraic sum of all these resultant displacements is thus shown by the heavy line, Fig. 387. In this way, a wave of perfectly definite length is produced, but differing in *form* from that of either of the component waves. When, therefore, *two waves of different lengths* interfere, they combine to produce a wave of different *form* from that of either of the component waves. This analysis will enable us the better to understand the production of beats, and

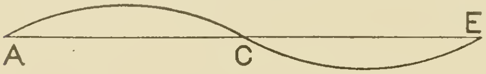

Fig. 385. A simple wave.

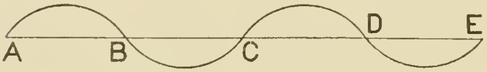

Fig. 386. A simple wave of half the above wave length.

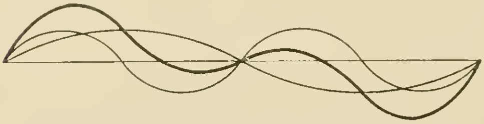

Fig. 387. Resultant wave formed by the composition of the above two simple waves.

will also serve to show why it is that when two notes are sounded together, the resultant tone produces a different sensation upon the ear from that produced by either of the component notes when sounded separately.

575. The Form of the Resultant Wave. The resultant wave, produced by the composition of two simple waves, may be most beautifully shown by arranging two tuning forks so that one carries a piece of smoked glass attached to one of its prongs, while the other is provided with a stylus, likewise mounted upon one of its prongs. If the two forks are now made to vibrate, and the one with the stylus is drawn over the other, the curve which is traced upon the smoked glass will be the resultant curve produced by the composition of the two simple harmonic curves which each fork would have traced alone. If the two forks make precisely the same number of vibrations per second, the resulting curve will be a simple sine wave, the amplitude depending upon the phase in which the forks are started. If, however, one fork vibrates slightly faster than the other, as, for instance, if one makes 81 vibrations per second while the other makes only 80, one fork gains one vibration per second over the other. If, then, at the beginning of the second the two are vibrating together, or are in the same phase, at the end of the second they will again vibrate in the same phase, but in the middle of the second they will vibrate in opposite phase. When the two are vibrating in the same phase, the amplitude of the resultant wave is the sum of their separate amplitudes, and when they are vibrating in opposite phase the resultant amplitude is the difference between these amplitudes.

The resultant wave, therefore, is far more complex than either of the component simple waves, and consists, in this particular instance, of 81 crests and troughs of different amplitudes, all of which must be executed before the wave length of the resultant wave is complete. Once a second, therefore, the intensity of the sound increases, and once a second it decreases, thereby producing a throbbing, or pulsating, effect, which is known as the phenomenon of beats. The variations in the intensity of the sound when beats are heard are thus seen to be caused by the inequalities in the amplitude of the resultant wave, and by the long periods of time occurring between these inequalities, thereby enabling the ear to detect the variations in the intensity of the sound thus produced. This arrangement of the two tuning forks, which thus makes possible the actual construction of the form of the resultant wave, has well been called the optical method of studying beats.

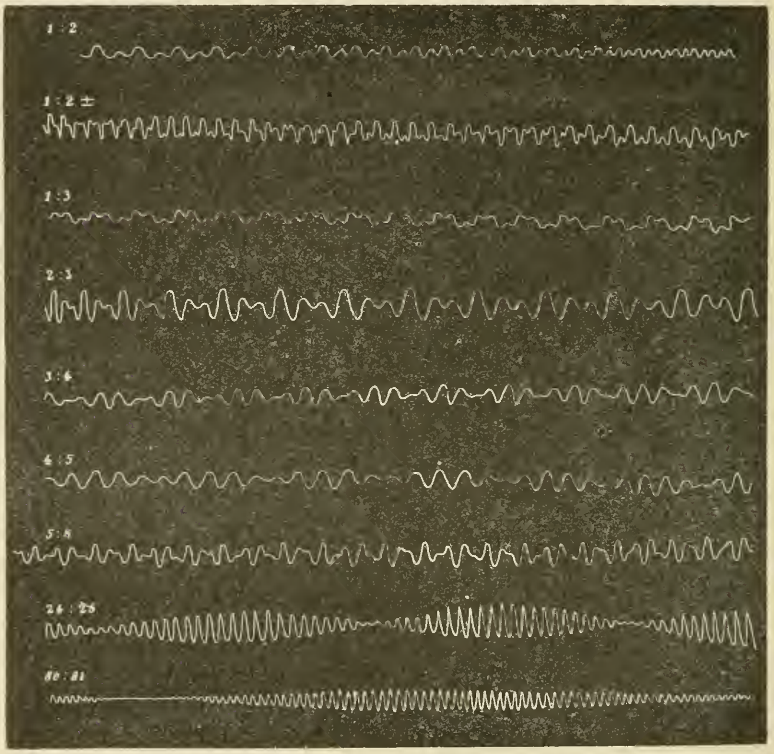

Fig. 388. Resultant waves formed by the composition of two simple waves of different wave lengths.

If the interval between the two notes is smaller, the more quickly do the variations of amplitude which determine one wave length repeat themselves in the one following, and the shorter is the wave length of the resultant vibration. Accordingly, *the interval between the two notes determines the length of the resultant wave;* and the *form* of this wave becomes *less and less complex* as this interval is expressed by *smaller and smaller* *numbers.* When the interval can be represented only by large numbers, such as 81/80, the resultant wave, which is completed only after many vibrations of the component forks, is therefore exceedingly complex, and the greatest possible discord, namely, that produced by beats, is heard. Even when the interval is expressed by numbers as small as 9/8, which occurs when C and D are sounded together, the resultant wave is completed after nine vibrations of one of the component forks; and although this wave is a far simpler vibration than the one just cited, it is nevertheless so complex as to produce the sensation of a discord.

When, however, the notes E and G, which form the interval 6/5, are sounded together, the resultant wave is completed after six vibrations of one of the forks; and while this wave is still far from simple, it is less complex than others having an interval expressed by larger numbers, and the ear recognizes these two notes as tolerably accordant, although the smoothness of the sound is not nearly so complete as that perceived with notes having intervals expressed by still smaller numbers. Fig. 388 gives an idea of the form of the resultant waves formed by the combination of two notes of different intervals. It will be seen in every instance that the accord becomes more nearly complete as the form of the resultant wave gains in simplicity, while the discord becomes more apparent as the resultant wave increases in complexity.

576. Beat Tones. Beats have been defined as the periodic variations in the intensity of a sound produced when two notes which do not differ widely from unison are sounded together. As the notes differ more and more from unison, the beats come with greater and greater rapidity, and these, like any other periodic vibrations, may be detected as individual sounds until their number becomes about 30 per second. If the beats occur more rapidly than 30 per second, they blend together into a musical note, the pitch of which depends upon their number per second. The tone thus produced by the blending together of beats is called the *beat tone.* For example, if the notes C and G are sounded together, C making 256 vibrations per second and G 384, there will then occur 128 beats per second, and these will blend together into a note, or beat tone, having 128 vibrations per second, and therefore of a pitch an octave lower than the C. In sounding two notes together, this third or beat tone is always to be considered; and, in general, two notes when sounded together can produce a pleasing sensation only when the beat tone harmonizes with each of the components, that is, when the vibration-frequency of this beat tone bears a simple ratio to that of each of the original notes.

577. Timbre or Quality. Whenever the same note is played upon two musical instruments, the ear detects a certain difference in the sounds of these notes, although both are of the same pitch and both may be equally loud. This difference, which thus characterizes each note as peculiar to the particular instrument upon which it is played, is known as the *timbre,* or the *quality,* of a musical note. A careful investigation of the vibrations of a sounding body reveals the fact that no source of sound vibrates with a perfectly simple vibration, but that every vibrating body to a greater or less degree subdivides into a series of vibrating segments, even while it may also be vibrating as a whole. There are always, therefore, superposed upon the fundamental, which gives the pitch to the tone, also a number of vibrating segments of the fundamental, and these, accordingly, form overtones or harmonics. In other words, a sounding body may vibrate with its fundamental, while at the same time it may vibrate also with a certain series of harmonics which exist upon this body just as if the fundamental were not present, Figs. 389 and 390.

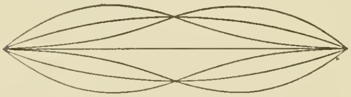

Fig. 389. The first overtone or harmonic vibrating in conjunction with the fundamental.

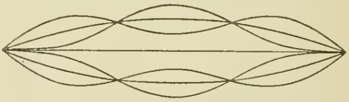

Fig. 390. The second overtone or harmonic vibrating in conjunction with the fundamental.

Upon every vibrating body, therefore, there exists a number of independent vibrations of different periods, consisting of the fundamental and a certain number of segments, and these interfere to produce a resultant vibration which is far from simple. Each component vibration originates waves of different lengths in the air, and these interfere to produce a wave which is the resultant of all the component waves. The actual form of a sound wave, even from a single note of any instrument, is therefore really very complex, having superposed upon the fundamental a number of overtones which are peculiar to that instrument. The construction and resonance of any instrument will cause one set of overtones to be formed or especially re-enforced, while the different form or even the different handling of this instrument will bring into prominence a different series of overtones or harmonics in conjunction with the fundamental. It is, then, the presence of certain overtones which are superposed upon the fundamental that gives the particular character to the tone played upon one instrument, while another set of overtones occurring with the fundamental will cause the same note upon another instrument to have an entirely different character. The *timbre*, or the *quality*, of a musical note, therefore, *depends upon the particular overtones which occur in conjunction with the fundamental and upon their intensity.*

Thus, a flute, a horn, a piano, and a violin may all sound the same note and may be recognized as being in perfect accord, and yet there are wide differences in the *quality* of the sound. This means that each is vibrating with a fundamental of the same period, and this fundamental gives the predominant pitch to the note; but superposed upon this fundamental are certain overtones in one instrument, and different overtones in another. The squeaky, rosiny sound emitted by the violin indicates that the violin is peculiarly rich in the higher overtones; and these particular overtones, which depend upon the form and upon every item of the construction of the instrument, when combined with the fundamental, give the violin note its special characteristic, or *quality*.

It is a singular fact that, certainly as regards any difference which the ear can detect, the *phase* in which these harmonics occur *is without influence in affecting the timbre, or the quality, of the sound produced.* This is due to the fact that the ear resolves a complex vibration into its fundamental simple ones, and these, by the frequency with which they occur and by their intensity, are alone effective in modifying the sensation produced by the fundamental.

A note upon the piano may be vibrating with one or more of the harmonics superposed upon the fundamental, although we may be conscious only of the fundamental. This may be shown, and at the same time the phenomenon of sympathetic vibrations may be nicely illustrated, by gently pressing down any key so as to lift the damper without causing the string to sound, and then striking violently the note an octave below. If the vibrations of the lower note are then stopped, the note which is held down will be heard to sound faintly, having been set into sympathetic vibration by the vibration of the lower note. But since one body sets another in sympathetic vibration only when the periods of the two are the same, we see that the lower note must have been vibrating with a period equal to that of the upper one, or, in other words, the first overtone was combined with the fundamental in the lower tone, and the upper note vibrated sympathetically with this overtone.

That a stretched string is peculiarly rich in harmonics combined with the fundamental may be shown by picking the string of a sonometer, then touching it lightly with the finger at the center, thus stopping the fundamental vibration, when the first overtone, formed by the division of the string into two segments, will be clearly heard, Fig. 389. In the same way, if the string is touched at a point one-third of its length from the end, Fig. 390, one-quarter of its length from the end, etc., the second, the third, the fourth overtones are heard, and so on up to the tenth, or even higher. This shows that upon the fundamental of a string there are superposed all the harmonics up to the tenth and even beyond, and that the peculiar *twang* of a string is the resultant tone produced by the fundamental in conjunction with all of these harmonics.

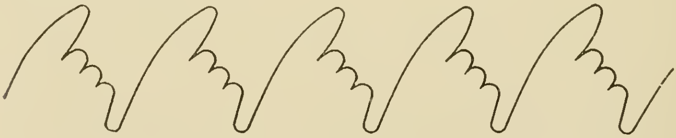

Fig. 391. The complex wave produced by the coexistence of harmonics and fundamental upon a vibrating string.

The resultant complex form of the wave, produced by the superposition of these various wave motions, may actually be seen when a small portion of the vibrating string is projected upon the screen by a revolving lens. If the lens were stationary, the string would be seen to vibrate, but these vibrations would occur so rapidly that their character could not be studied. The moving lens, however, carries across the screen the image of a section of the vibrating string, thus forming a wave, Fig. 391, the form of which is seen to be very complex.

578. The Composition of Two Harmonic Vibrations at Right Angles to Each Other. In the treatment of the composition of wave motions which we have heretofore considered, we have assumed that the vibrations of the particles take place parallel to one another, and that the motions of the waves are in the same direction. Lissajous conceived the idea of studying resultant vibrations by compounding two simple vibrations taking place at right angles to each other, and obtained results of singular beauty and interest. Before discussing these results, as well as Lissajous's mode of procedure, we must turn our attention to the graphical composition of two harmonic motions when these take place along paths at right angles to each other.

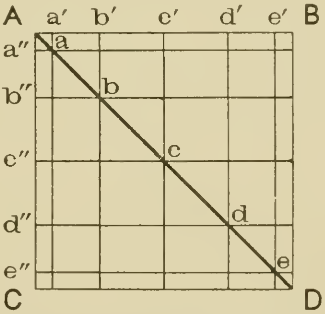

Fig. 392. Graphical composition of two harmonic motions at right angles to each other.

Let us assume that a particle is actuated by a harmonic motion in the horizontal direction along the line AB, Fig. 392, and at the same time also by a harmonic motion in the vertical direction along the line AC. We will assume, furthermore, that the two vibrations have the same period and amplitude, and that they occur in the same phase, that is, that the particle is at its greatest upward displacement due to the vertical motion at the same instant that it is at its greatest displacement to the left due to the horizontal motion. In other words, we will assume the particle to be at the point A, and just about to begin its vibration along both directions. If, now, for convenience, we think of the period divided into twelve equal parts, we may then divide the paths AB and AC into six harmonic divisions, which will represent the distances in each direction which the particle travels in equal times, and then, by the composition of motions, we may determine the position of the particle at each twelfth of a period.

Since the horizontal and the vertical motions begin at the same instant, the particle, after leaving *A*, will have reached the point a' in the horizontal direction at the same instant that it reaches the point a'' in the vertical direction; and, therefore, after one-twelfth of a period the particle will be found at the point *a*, having actually moved along the diagonal *Aa* of this little parallelogram. In the same way, the particle will appear successively at *b*, then at *c*, and so on; hence in one-half of the entire period, the particle will have moved along the line *AD*, and during the second half period it will retrace this path in the opposite direction. The particle, therefore, will move with simple harmonic motion along the straight line from *A* to *D* and back again. Any simple harmonic motion, therefore, may be regarded as the composition of two simple harmonic motions in the same phase and of equal periods and amplitudes at right angles to each other.

If the two motions are in slightly different phase, as, for instance, if the horizontal vibration is one-twelfth of a period in advance of the vertical, then the particle is at a' when the vertical motion downward begins, and the flattened ellipse, shown in the second diagram, Fig. 393, is the curve traced by this particle. If the phase of the horizontal vibration differs from that of the vertical by one-sixth of a period, the particle will have reached the point b' when the vertical motion begins, hence the particle will trace the open ellipse, shown in the third diagram, Fig. 393. If the horizontal vibration is in advance of the vertical vibration by one-quarter phase, the particle will be at c' in its horizontal motion when its vertical motion begins, and

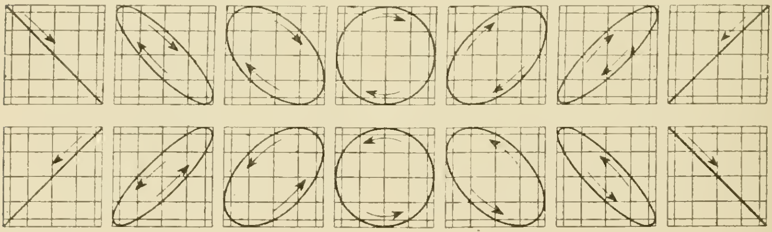

FIG. 393. Composition of two harmonic motions of equal period and amplitude at right angles to each other. In each of the diagrams the phase of the horizontal vibration is advanced one-twelfth of a period over the preceding.

the path traced will then be the circle, shown in the fourth diagram, Fig. 393. Motion in a circle, therefore, may be regarded as the resultant of two simple harmonic motions at right angles to each other, of equal period and amplitude, but differing in phase by one-quarter period. The succeeding figures show the variations in the forms of these curves produced in each case by increasing the phase of the horizontal vibration by one-twelfth of the period over that of the verical.

If we now assume, as before, that the two component vibrations are of equal amplitude, but that the period of the horizontal is only one-half that of the vertical, the particle will move twice as rapidly in the horizontal direction as in the vertical, travelling over two spaces in the former direction, while it moves over but one in the latter. If, now, the two vibrations start in the same phase, that is, both leaving the upper left-hand corner of the figure together, the resultant curve traced is shown in the first diagram, Fig. 394. Similarly, if the horizontal vibration is one-twelfth of the period in advance of the vertical, the resultant curve is shown in the second diagram, Fig. 394. In the same way, the succeeding figures show the resultant curve traced in each case when the horizontal vibration is advanced one-twelfth of the period over the preceding one. Curves of exceeding complexity result when the interval between the component vibrations is greater than the simple ones here illustrated.

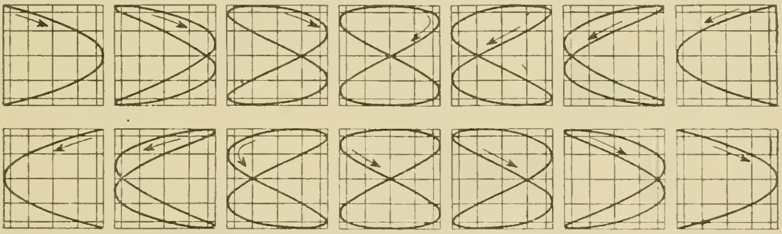

FIG. 394. Composition of two harmonic motions of equal amplitude at right angles to each other, the period of the horizontal being one-half that of the vertical. In each diagram the phase of the horizontal vibration is advanced one-twelfth of a period over the preceding.

579. Lissajous's Figures. In order to obtain these results experimentally, Lissajous mounted two tuning forks at right angles to each other, and placed a small mirror upon a prong of each. The two forks were then so arranged that a beam of light could be thrown upon one of these mirrors and then reflected to the other, Fig. 395, after which it was reflected upon a screen. Since the forks are at right angles, if one fork vibrates, the vertical one for instance, while the other remains stationary, the spot of light upon the screen will vibrate vertically with harmonic motion. If, on the other hand, the horizontal fork is made to vibrate while the vertical one remains stationary, the spot of light on the screen will vibrate horizontally with harmonic motion. If, now, both forks are made to vibrate at the same time, the spot of light will move vertically with harmonic motion due to the vibration of the first fork, and horizontally with harmonic motion due to the vibration of the second, and hence will move with a motion which is the resultant of these two. The fact that the eye retains for some time an impression made upon it causes this moving spot of light to trace upon the screen a curve which, by the persistence of vision, seems continuous. The form of this curve depends upon the amplitudes and periods of the two component motions, and upon the particular

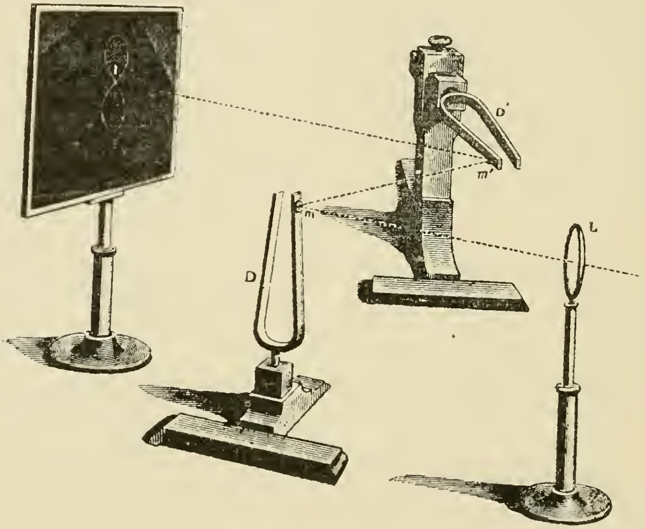

Fig. 395. The formation of Lissajou's figures. This cut is reproduced from Zahm's "Sound and Music" through the courtesy of A. C. McClurg & Co.

phase in which the two vibrations happen to be started. If the periods of the two forks do not bear a simple ratio to each other, the phase of the resultant vibration will change, and the figure on the screen will be seen slowly to pass through all the variations in the forms of these curves which can be produced by a change of phase.

LIGHT.

CHAPTER LII.

THE NATURE, VELOCITY AND INTENSITY OF LIGHT.

580. Methods of Treatment. Light is the phenomenon of the vibration of the ether which produces the sensation of sight. The study of light, like the treatment of sound, may be considered either subjectively or objectively. In the subjective treatment, we discuss the arrangement and properties of the different media of the eye, the structure of the retina, and the processes by which an impression which is received upon this sensitive membrane is transmitted through the optic nerve to the brain. In the objective treatment, we have to deal with the action which is taking place in the luminous body, the communication of this action, or disturbance, to the surrounding medium, and the method by which this disturbance is finally propagated to the eye. It is mainly with the latter, or the objective treatment of light, that we shall have to deal in the following pages.

581. The Rectilinear Propagation of Light. One of the simplest phenomena connected with the subject of light is the fact that light travels in straight lines. If an opaque object is interposed between a point source of light and a screen, the straight lines drawn from the light to the edges of the object, Fig. 396 mark the boundaries of the geometrical shadow. The space outside the lines thus drawn is illuminated, while the region inside is shaded. If the source of light is extended in area, Fig. 397, the boundaries of the shadow are not sharp. This is due to the fact that, while light from one portion of the source may be intercepted by the opaque object, light from another portion may pass by the edges of this object and so produce an illumination. The shadow upon the screen, then, really consists of two parts, one, a totally shaded area, called the *umbra*, from which the light is entirely excluded; and the other, a surrounding, partially darkened region, called the *penumbra*, which receives light from a portion only of the source, and which varies gradually in intensity from the total obscurity of the umbra to the full illumination of the unshaded screen. If, however, the source of light is small, the shadow is very sharply marked, thus showing that light bends to a very slight extent indeed into the space included within the geometrical shadow. Under ordinary conditions, therefore, light does not turn a corner.

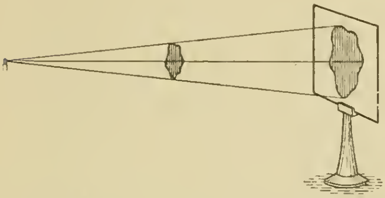

FIG. 396. If the source of light is small, the shadow is sharply defined.

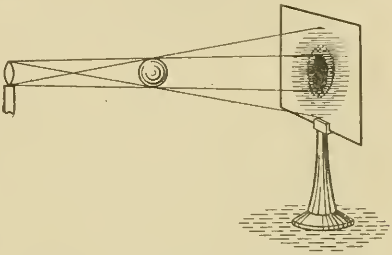

FIG. 397. If the source of light is large, a penumbra surrounds the umbra.

If a hole is made in an opaque screen placed before a source of light, rays of light from every part of the source pass in straight lines through the opening and as they fall upon a second screen placed beyond, Fig. 398, they produce an inverted image of the source. Since each part of the image is connected with the corresponding part of the object by a straight line, the size of the image will be greater the farther the second screen is removed from the first, although the brightness of this image decreases as its size increases.

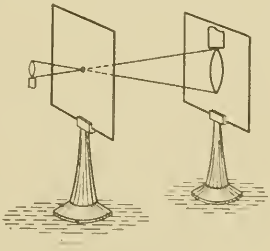

FIG. 398. Image produced by a small aperture, showing the rectilinear propagation of light.

A *ray* of light may be defined as the direction in which light is propagated. It has no real existence, but is merely a *direction* determined by the effects of screens and apertures.

582. Theories as to the Nature of Light. If we enter an unlighted room at night we are entirely unconscious of the contents of the room until a light is brought, after which the various objects at once become visible. It is evident, therefore, that the possibility of perception by the sense of sight depends primarly upon a source of light by which objects are illuminated, and that something must proceed from this light as a center and ultimately reach the eye, in order that the sensation of sight may be produced. What the nature of this cause of light may be is a matter which has been studied with great interest by men of science for many years. As early as the middle of the seventeenth century there were already formulated two theories to account for the phenomenon of light. One of these, known as the emissive or corpuscular theory, was warmly supported by Newton, while the other, known as the wave theory, was propounded by the great mathematician Huygens.

According to the emissive or corpuscular theory, it was thought that little particles or corpuscles of light were continually projected through space in straight lines with an enormous velocity, and that these little missiles or projectiles, when they impinged upon the retina of the eye, produced the sensation of sight. The main reason for the general adoption of this theory was the fact which was then well understood, that light travels in straight lines. Sound was thought to be a wave motion because it can turn a corner and pass behind an obstacle; but since light travels in straight lines and does not to any great extent pass into the geometrical shadow, it was argued that light cannot be a wave motion. Moreover, the fact that light shadows are sharply defined was perfectly consistent with the emissive theory, because it was thought that sharp shadows could be produced only by the fact that the opaque screen intercepted the little corpuscles of light which travelled in straight lines from the source. This, which was the simplest and most plausible theory, was the one first adopted by the earlier investigators, and at the time of Newton it was the one that was generally accepted. Indeed, Newton himself advocated so earnestly the corpuscular theory, that the wave theory of light, which Huygens, his contemporary, proposed, received for a century and a half but little attention.

In the year 1801, however, Thomas Young pointed out a series of phenomena which were wholly at variance with the emissive theory, but which admitted of ready explanation upon the wave theory. Although meeting at first with violent opposition, this test proved to be crucial and final; and since that time it has been recognized not only that the corpuscular or emissive theory cannot be true, but also that the wave theory as proposed by Huygens is undoubtedly the one which correctly interprets the various phenomena of light.

583. The Electromagnetic Theory of Light. The adoption of the wave theory necessitated the assumption that the luminous body, or a portion of it, was in a state of vibration, and that the vibrations thus produced were the origin of the light waves. Inasmuch as the light emitted by a substance in the gaseous form is characteristic of that substance, it was thought that the vibrating particles must be the atoms, since these alone were known to be different in different bodies. It is true that Maxwell, in 1873, as the result of a profound mathematical investigation, showed that the propagation of light might be explained as taking place by electromagnetic waves in the ether of space, but this left still undecided the nature of the vibrations in which the ether waves found their origin.

It was, therefore, not until the discovery of the electrons, and the undoubted proof that these particles exist not only among the molecules of matter but also in orbital motion within the atoms themselves, that a cause was found for the production of the electromagnetic waves demanded by Maxwell's theory. Since a magnetic as well as an electrostatic field accompanies a moving electron, the ether in the neighborhood of this particle is constantly under strain. Accordingly, when the electron is *accelerated,* that is, when it changes either its velocity or its direction of motion, the strain or distortion of the ether changes also, and an electromagnetic wave goes out from this particle as a center, and is propagated through space with the velocity of light. From this it follows that when the free electrons and the molecules of matter, with which they are in thermal equilibrium, mutually bombard one another, they are constantly accelerated, since their velocities as well as their directions of motion are continually changing.

The accelerated electrons may therefore be regarded as the ultimate source of the ether waves; and when these ether waves break upon the eye with the requisite rapidity and intensity, they stimulate the retina in such a way as to produce the sensation of light. This hypothesis, which, indeed, is only a modification of Huygens's wave theory, enables us to form a clearer conception of the cause and nature of light than has ever before been possible.

584. The Ether. The hypothesis that light is a wave motion demands the further assumption that there must be a medium filling all space, the vibrations of which produce the sensation of light. Since light travels through a vacuum and also through the interstellar spaces, which are wholly devoid of air, it was soon seen that the atmosphere cannot be the medium by means of which light is propagated. It has, therefore, been necessary to assume the existence of another medium called the ether, extending throughout all space even to the remotest fixed stars; and the vibrations of this ether, when they fall upon the retina of the eye, are believed to produce the impression of sight. The only direct evidence of the existence of the ether is the fact that a luminous body originates waves in this medium, and that the wave length, the number of vibrations per second, the direction of vibration, and the velocity of propagation are all capable of measurement. The almost inconceivably great velocity with which light, or this wave motion in the ether, is propagated, points to an enormously high elasticity of the ether, and to a density so small, when compared with other material, ponderable bodies, that it cannot in any way be measured or even detected. Numerous experiments, which will be taken up in detail later, show conclusively that the vibrations of this ether are transverse. This can be interpreted only as meaning that the ether has elasticity of form, and therefore that this medium must be of the nature of a solid. It is on this account that the wave theory of light is sometimes called the elastic solid theory.

585. Huygens's Construction for the Propagation of a Spherical Wave. Following Huygens's assumption that light is a wave motion, we must think of the electrons in a luminous body as being in a state of vibration, and of these vibrations as being communicated in all directions to the surrounding ether. Each particle of the ether, therefore, is thrown into vibrations precisely similar to those of the electrons in the luminous body which constitute the source of the light waves. Moreover, since the ether is homogeneous, or alike in all directions, the vibratory motion is propagated with equal rapidity in all directions, so that after the lapse of a certain time, the disturbance is found upon the surface of a sphere which has the luminous body as the center. The spherical wave of light, or a portion of this spherical surface, is known as the *wave front.* If the source of light is very distant, so that the radius of this spherical wave is very great, a small portion of this sphere may be considered plane, and we may then speak of a *plane wave front.*

Since light is propagated with equal velocity in all directions, all points upon the wave front are vibrating in the same phase. *The wave front, therefore, may be defined as the continuous locus of all points in the same phase of vibration.* But if the particles of the wave front are executing vibrations identical in character with those of the electrons in the luminous body, each particle upon the wave front must be considered as a center of disturbance sending out little spherical waves precisely similar to those at the center. Little elementary waves, therefore, start out from each particle on the wave front as a center, Fig. 399, and the envelope of all these little elementary waves is the new wave front. This process is then repeated. As each particle of this new wave front is thrown into vibration, it becomes itself a center of disturbance which sends out little ele-

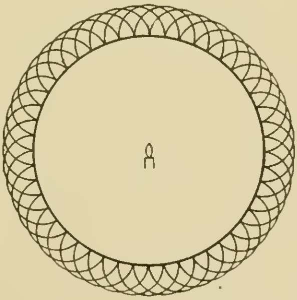

FIG. 399. The propagation of a spherical wave due to the Huygens wavelets.

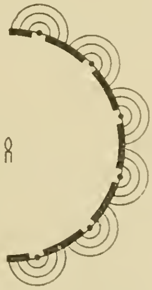

FIG. 400. Huygens wavelets, originating from the unobstructed portions of a wave front.

mentary waves, and the envelope of all these little elementary wave fronts becomes, in turn, the new wave front. Thus it is that the spherical wave advances from one position to another.

Inasmuch as every particle of the wave front becomes a center of disturbance which sends out little elementary waves, the envelope of these latter waves is continuous, and the resulting wave front is perfectly smooth and spherical. If, however, portions of the wave front are stopped by small screens, Fig. 400, the wavelets then emanate from the spaces between the screens as centers; but in this case these centers are far enough apart to enable the existence of the little wavelets to be detected, and the wave front is then found to be no longer spherical and continuous. The same phenomenon, however, would have taken place, and the same series of wavelets would have been formed, if the little screens had not intercepted certain portions of the wave front. The fact, therefore, that an unobstructed wave front advances in the spherical form, is evidence that *each* vibrating particle upon this surface becomes a center of disturbance which sends out elementary waves, and the envelope of these little wave fronts becomes the new wave front.

586. Wave Fronts of Light are Analogous to Sound Waves. While the presence of spherical wave fronts of light in the ether cannot be experimentally demonstrated, the fact that such wave fronts do exist and are propagated in the method just described may be beautifully shown in the analogous case of sound waves in the air.

The variations in the density of the air produced by the condensations and rarefactions of waves of sound cause a very minute bending of rays of light which pass through such waves. By allowing these rays to fall upon a sensitive plate, Wood has been able to photograph the forms of sound waves as they exist in the air. In these experiments, the source of disturbance which gives rise to waves of sound is the sharp snap of an electric spark, produced by the discharge of a Leyden jar. A wave of aerial condensation and rarefaction, starting from this spark as a center, Fig. 401, advances in a spherical form through the air. About the ten-thousandth of a second later, another distant spark is produced, giving a light which illuminates for an instant the spherical wave of sound produced by the first spark, and this light, falling upon a photographic plate, causes a shadow of the sound waves to be recorded. Although the details of the operation are not simple, the process, in brief, may be described as taking a flashlight picture of a sound wave after it has advanced a short distance from its source of disturbance.

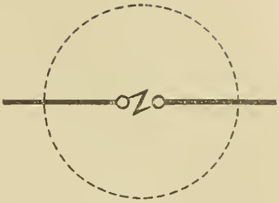

FIG. 401. The snap of an electric spark, giving rise to a sound wave.

By placing little screens where they will intercept small portions of the spherical wave front, the effect produced by the unobstructed portion of the wave front may be studied. It is then seen, Fig. 402, that small spherical wavelets originate from the spaces between the screens as centers, and pass out as little spherical waves into the space behind the screens. This same phenomenon would have taken place with this portion of the front had the little screens not been present; and therefore it is seen that every particle of the medium of the wave front is a center of disturbance which sends out spherical waves precisely similiar to those at the center, and that the new wave front is simply the envelope of all these elementary spherical waves.

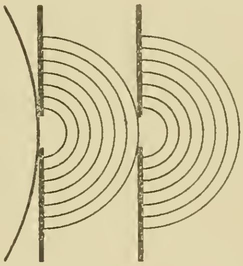

FIG. 402. Elementary sound waves proceeding from vibrating particles of the wave front.

587. The Velocity of Light. Since light travels with a velocity almost inconceivably great, all early attempts to measure the time required for light to pass from one place to another upon the earth's surface were without avail. It was not until the discovery of the telescope, thus enabling interplanetary spaces to be employed, that it was possible to detect the finite time required for light to travel over a certain distance, and thus to determine its velocity.

588. Roemer's Determination of the Velocity of Light. In the year 1676, Roemer, a Danish astronomer who was then living in Paris, was engaged in studying the innermost of the four principal satellites of the planet Jupiter. He observed that the plane of the orbit in which this moon revolves practically coincides with the plane of the earth's orbit. To an observer on the earth, therefore, this little satellite, which really revolves in a circle about its primary, appears to vibrate with harmonic motion along the diameter of its orbit (p. 56), passing alternately across the face of the planet towards the earth, and behind the planet away from the earth and from the sun. When on the farther side, the little moon enters the cone of the shadow cast by the planet, and so is eclipsed.

As long as the earth was in the part of its orbit nearest to Jupiter, Roemer observed that the satellite suffered eclipse at perfectly regular intervals. From the period of this body's revolution, which was thus determined, the times at which future eclipses would occur could be predicated with great accuracy. It was soon discovered, however, that as the earth receded from Jupiter, the interval between eclipses was greater than the known period at which these occurred, and that when the earth again approached the planet, these intervals became shorter than this period. When the earth was on the opposite side of the sun and farthest from Jupiter, Fig. 403, the eclipses took place once more with their former regularity, although each eclipse was found to begin 996 seconds later than the time which had been predicted six months before. (This is the value for the interval which is known to be correct. The computed value, as Roemer determined it, was somewhat greater).

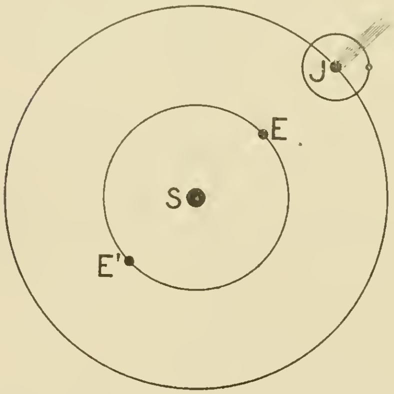

Fig. 403. The earth in two positions with reference to the sun and to Jupiter.

Roemer saw that all of these results were consistent with the hypothesis that light travels with a finite velocity. It was obvious that the eclipses must take place in uniform periods, and would be observed to occur at regular intervals upon a stationary earth. While the earth was receding from Jupiter, the light from each eclipse had a farther distance to travel than that from the eclipse preceding, and, similarly, when the earth was approaching the planet, the light from each eclipse had a correspondingly shorter distance to travel. All of these discrepancies for the period of recession or approach amounted to 16 minutes and 36 seconds, and the aggregate increases of distance which the light had to travel when the earth was farthest from Jupiter, was then the diameter of the earth's orbit, or 186,000,000 miles. This gives as the velocity of light approximately 186,000 miles per second, which is very nearly the accepted value at the present time. Although Roemer's attribution of these effects to the finite velocity of light was at first discredited, it has since that time been corroborated in the most satisfactory manner, and the velocity with which light travels is now known to differ not widely from 300,000 kilometers per second.

589. The Aberration of Light. A second determination of the velocity of light by the employment of astronomical means was made by Bradley in 1726. Bradley discovered that the apparent positions of the fixed stars vary with the motion of the earth, so that the stars all seem to move one way during one-half of the revolution of the earth about the sun, and in the opposite direction during the other half of this revolution. An investigation as to the cause of this apparent displacement of the stars from their true positions led to the discovery of the phenomenon of the aberration of light.

The principle upon which this effect depends may be illustrated by imagining a person to carry an umbrella in a rain which comes vertically downward. As long as the person stands still, the umbrella shields him completely, and he perceives the rain falling in its true direction. If, however, he moves forward, still carrying the umbrella in the same position, he may move so rapidly as to intercept a drop of rain which passes the edge of his umbrella, and the rain then strikes his face. *To the person,* therefore, the rain *apparently* comes in an inclined direction, which is really the resultant of his own motion forward and the motion of the rain downward. Moreover, to shield himself entirely from the rain, he will have to incline his umbrella in the direction in which he is moving, the amount of this inclination being greater as his own motion forward increases.

In an analogous manner, the phenomenon of the aberration of light depends upon the relative velocities of the earth and of light. Thus, if *S*, Fig. 404, represents a star directly overhead, and *AB* is a telescope through which the star is observed, the light from the star, falling upon the telescope at *A*, would pass down the tube to the observer's eye at *B*, if the telescope were upon a stationary earth. But during the time required for the light to travel the length of the telescope from *A* to *B*, the telescope has moved from *B* to *B′*, being carried forward by the earth in its motion of revolution around the sun, and the light, instead of reaching the observer's eye at *B*, will strike against the side of the tube. The tube of the telescope, therefore, must be inclined forward in the position *AB″* by such an amount that the eye-piece *B″* of the telescope will be in the position *B* when the light which enters *A* reaches the bottom of the tube. The light, therefore, in reality travels along the straight line *AB*, although to the observer stationed at *B″* it *appears* to come in the direction *AB″*.

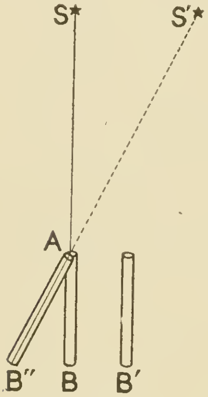

FIG. 404. The displacement of a star due to the aberration of light.

Inasmuch as an object is always seen in the direction of the ray of light which enters the eye, it follows that, to the observer at *B″* the star does not seem to be in its true position at *S*, but appears to be displaced to the position *S′* that is, in the direction of the motion of the earth. The amount of this displacement depends upon the direction in which the earth moves with reference to the star, and reaches the maximum value of 20″.5 when the motion of the earth and the direction of the star are at right angles to each other. Since the distance *B″B*, is traversed by the earth in the time required for the light to pass from *A* to *B*, the ratio of these distances is also the ratio of the velocity v of the earth in its orbit to the velocity V of light. Accordingly, we have the relation

$$\frac{v}{V} = \tan \alpha,$$

where α is the above maximum angular displacement of a star, and is known as the angle of aberration. The tangent of $20''.5$ is about 1/10,000, and the known velocity of the earth in its orbit is approximately 30 kilometers per second. From this, the value of the velocity of light is determined to be not far from 300,000 kilometers per second.

The displacement of the stars from their true positions, due to the motion of the earth, is known as the *aberration of light*, and has been interpreted as pointing to a relative motion between the earth and the ether. When, however, experiments involving these quantities are tried in the laboratory, no such relative motion has been found to exist. The ether then appears to be at rest with respect to the earth, as if the latter, in its motion through space, carried with it the ether in its immediate neighborhood. Nevertheless, the two experiments are hardly identical in nature, inasmuch as the light from the star proceeds from a fixed source, while that from the terrestrial luminant necessarily partakes of the motion of the earth. This entire subject has, of late, given rise to much discussion; and it must be admitted that the true motion of the earth relative to the ether, and the extent to which the ether is dragged along by the earth in its passage through this medium, are problems which at present are not well understood.

590. Fizeau's Determination of the Velocity of Light. In 1849, Fizeau measured the interval of time required for light to travel from one point to another on the surface of the earth, and thus determined the velocity of light solely by terrestrial measurements. The device which Fizeau employed is shown diagrammatically in Fig. 405. A ray of light from a source L passed between the teeth of a cog-wheel W, and was reflected by a distant mirror M directly back upon its own path, so that it again passed through the open space between the same two teeth. When the wheel was rotated with a moderate speed, the light could still be seen between the teeth as it was reflected from the distant mirror. At a much higher speed of rotation, it was found that, during the time required for the light to pass from *W* to *M* and back again to *W*, the wheel had turned by such amount that the light, which had passed between two teeth as it travelled from the source of light to the mirror, was now intercepted by a tooth as it was reflected back. The number of teeth in the wheel and the speed of rotation being known, the interval of time required for the wheel to turn from an opening to tooth was likewise known. But in this interval, the light had travelled from the wheel to the mirror and back again, and hence the velocity could be determined.

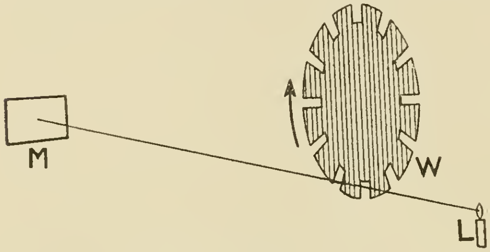

Fig. 405. Diagram of Fizeau's method of determining the velocity of light.

At a still higher speed of rotation, the light, which before had been eclipsed by a tooth on its return from the mirror, again appeared, thus showing that in the interval required for the light to pass from the wheel to the mirror and back again, the wheel had turned from one opening to the next. Accordingly, the light which passed between two teeth of the wheel as it started towards the mirror, passed between the next two as it returned, the wheel having rotated by this amount in the interval. Fizeau found a value for the velocity of light somewhat greater than 300,000 kilometers per second, but he himself felt that this was too great. Cornu repeated this experiment in 1874, and obtained a value almost identical with the results obtained by other methods.

591. Foucault's Method of Measuring the Velocity of Light. In the year 1850, Focault devised a means of measuring the velocity of light by a process not wholly unlike that of Fizeau; and his method, in the hands of later experimenters, has given results which are regarded as very nearly the actual value. The general plan which Foucault employed consisted in measuring the exceedingly short interval of time required for light to travel a measured distance on the surface of the earth.

Rays of light from a source *L*, Fig. 406, fell upon a mirror *M*, and were reflected to a distant mirror *N*, striking the latter normally, so that they were reflected back along the same path over which they had come. Accordingly, when the mirror *M* remained stationary, the ray of light *LM* was reflected to *N*, then back to *M*, and again to *L*. The mirror at *M*, however, was so mounted that it could be rapidly rotated; and in the interval of time required for the light to travel from *M* to *N* and back again to *M*, it turned through a small angle. The ray of light, therefore, when reflected from *N* to *M*, did not strike the mirror *M* at the same angle as that at which it left this mirror on its way to *N*, and accordingly, instead of being reflected to the original source *L* it was reflected to some new position such as *L′*. The angle *LML′* was then carefully measured; and since this angle was twice as great as that through which the mirror *M* had turned, as will be shown later, the latter angle, and therefore also the fractional part of a rotation which the mirror *M* had made, could be determined. But this fractional part of a rotation was the time required for the light to travel from *M* to *N* and back again to *M*. The distance and the time occupied by the light in traversing this distance being thus determined, the velocity of light could be at once computed.

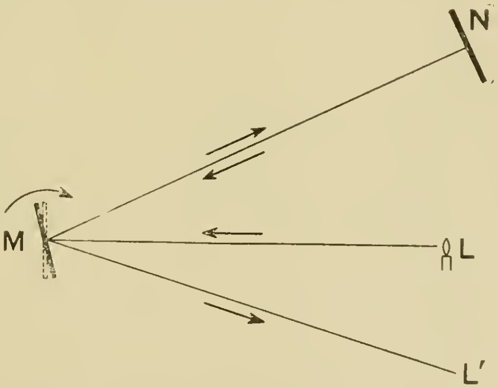

FIG. 406. Diagram of Foucault's method of determining the velocity of light.

Later, Michelson and Newcomb independently repeated Foucault's experiment, and obtained results agreeing almost exactly. The value for the velocity of light, resulting from the work of these investigators, may be taken as 299,860 kilometers per second, or, roughly 300,000 kilometers per second. It is extremely probable that this result does not differ greatly from the true value.

592. The Variation of the Intensity of Light with the Distance from the Source. Since we know that light is a wave motion of the ether, it is evident that, after the lapse of a certain interval of time, the disturbance, which originates at the source of light, is distributed over the surface of a sphere of area s and radius r, Fig. 407, where r is the distance the light travels during that interval. The entire surface of the spherical wave is in a state of vibration and represents a certain store of energy. After the lapse of another interval of time, the spherical wave has advanced in all directions, so that the energy, which was formerly distributed over the spherical surface s, having a radius r, is now found upon the spherical surface S, having a radius R. From this it follows that the amount of energy, and therefore, also, the amount of light falling upon unit area of each of these surfaces, must be inversely proportional to the extent of these surfaces.

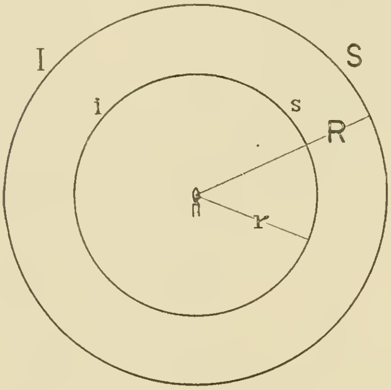

FIG. 407. Two spherical wave fronts about a luminous source.

If we now define *the intensity of light as the amount of light falling upon unit area,* it is evident that the relative intensities of the light, at the distances r and R from the source, are inversely proportional to the extent of the spherical surfaces s and S, having these distances as radii. Accordingly, if i represents the intensity of the light on the surface of the smaller sphere s, and I the intensity on the surface of the larger sphere S, then

$$i : I = \frac{1}{s} : \frac{1}{S}.$$

But

$$s : S = r^2 : R^2,$$

and therefore,

$$i : I = \frac{1}{r^2} : \frac{1}{R^2}.$$

Inasmuch as r and R are the distances of the respective surfaces from the source of illumination, it follows that *the intensity of light varies inversely as the square of the distance from the source.*

The truth of this law appears at once when we remember that the amount of light received by the area *a b c d,* Fig. 408 on the surface of the smaller sphere, is distributed over the area *A B C D* on the surface of the larger sphere. If, now, the radius of the larger sphere is twice that of the smaller, the area *A B C D* is four times that of *a b c d,* and the amount of light received *per unit area* on the larger sphere is only one-quarter of that received *per unit area* on the smaller. But the amount of light falling upon unit area is the measure of the intensity of the light; and therefore we see that as the distance from the source is made twice as great, the the intensity becomes only one-quarter as great, or that the intensity of light varies inversely as the square of the distance from the source.

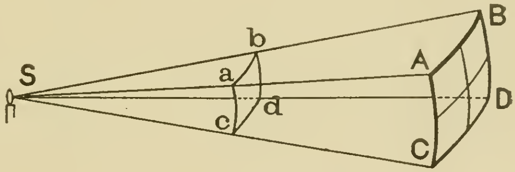

FIG. 408. The intensity of light varies inversely as the square of the distance from the source.

593. Photometry. In making photometric measurements, or determinations of the intensity of light, it is usual to compare the illumination produced by the light under examination with that from another source which is regarded as a standard. Since it is impossible directly to estimate the amount by which one light is brighter than another, the comparison between the two is accomplished by placing both before a screen, and adjusting their distances until the illumination produced by each is the same. *The intensities of the two sources are then directly as the squares of these distances.* This may be seen by considering that the two sources of light are the centers of two spherical waves in the ether which are tangent to each other at the screen, Figs. 409 and 410, and which have equal energy per unit area on their surfaces. The relative amounts of energy which these sources send out, or, in other words, the relative intensities of these sources, are proportional to the areas of these spherical surfaces, that is, to the squares of their radii, from which it follows that the intensities of the two sources vary directly as the squares of their respective distances from the screen.

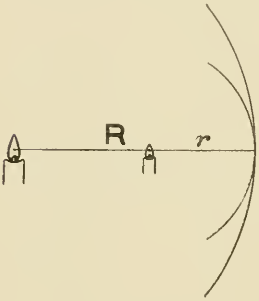

FIG. 409. Equality of illumination from two sources. Principle of the shadow photometer.

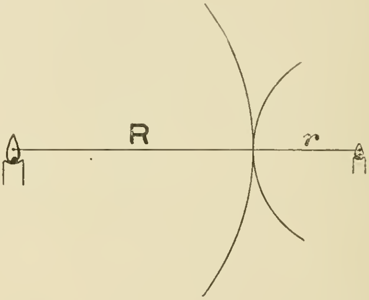

FIG. 410. Equality of illumination from two sources. Principle of the Bunsen photometer.

It is to be observed that the intensity of the *light* from a *single* source varies *inversely* as the square of the distance from that source. The relative intensities of *two sources*, however, vary *directly* as the squares of the respective distances of these sources from the point at which the illumination produced by each is the same.

In comparing luminous intensities, a unit of light called the candle power is used, the standard candle being made of pure sperm wax, weighing one-sixth of a pound, and burning 120 grains of wax per hour. The light, however, emitted even by the most carefully made candles is so exceedingly variable, that other standards have been tried, notably the light from a lamp which burns amylacetate, the wick being of a certain diameter, and the flame carefully adjusted to a certain height. This, too, while possibly more reliable than the standard candle, has proved unsatisfactory, and so photometric measurements may be regarded as only approximate.

594. The Shadow Photometer. Of the many forms of photometers, or devices for measuring light, which are in common use, two may be mentioned as the simplest and most practical. One of these, known as the shadow photometer, was devised by Count Rumford, and consists of a means of estimating the equality of the two shadows of an object, cast by the two sources of light to be compared. A standard candle *C* and the source of light *L* to be examined are placed near each other on a table, Fig. 411, and a rod *A*, or other opaque object, is so placed that it casts a shadow from each source upon a screen. The lights are then so adjusted that the two shadows cast by the rod are side by side and are equally dark. The shadow *M* is that of the rod cast by the standard candle *C*, but this shadow is illuminated by the light *L*; and the shadow *N* is that of the rod cast by the light *L*, but this is illuminated by *C*. The comparison of the uniformity of the two shadows, therefore, is simply a comparison of the equality of the illumination of these two shaded portions. The distance of each light from the screen is then measured, and the ratio of the intensities of the lights is the ratio of the squares of the distances.

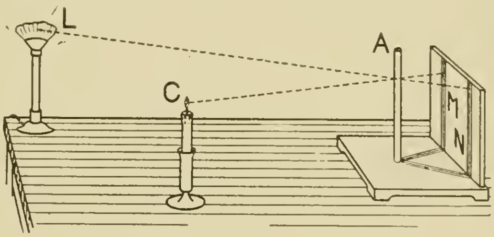

FIG. 411. The shadow photometer.

595. The Bunsen Photometer. Another device, known as the Bunsen photometer, consists of a long graduated bar, Fig. 412, with the standard candle and the source of light to be measured at opposite ends. The process of determining the relative

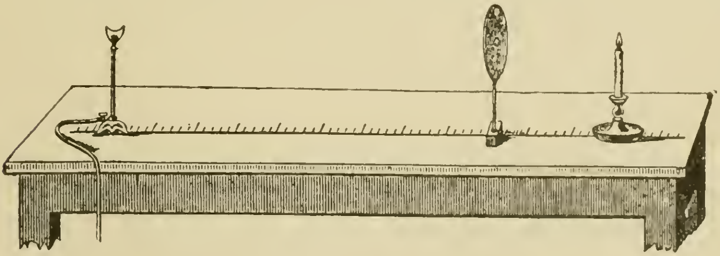

FIG. 412. The Bunsen photometer.

intensities of the two lights consists of finding the point on the bar where the illumination produced by one light is just equal to that produced by the other after which the distance of each light from that point may be read off from the scale beneath. The ratio of the squares of these distances then gives the relative intensities of the lights.

The great advantage of the Bunsen photometer consists in determining with great accuracy the point on the bar where the illuminations produced by the two sources are equal. The device by which this is accomplished consists of a sheet of paper with an oiled spot in the center, Fig. 413. This oiled spot appears darker than the paper by reflected light, but lighter than the paper by transmitted light. Accordingly, if the prepared piece of paper is placed between the two lights to be compared, it appears bright with a dark center when viewed from the side or the brighter light, and dark with a bright center when viewed from the side of the weaker light. When the illuminations on both sides are equal, the paper has a perfectly uniform appearance, and the oiled spot vanishes completely. The art, therefore, of "setting" the photometer consists of moving this screen of paper back and forth until the oiled spot in the center disappears, and the point at which this occurs is the position on the bar at which the illuminations produced by the two lights are equal. The distance from the screen to each of the lights is then determined, and the squares of the distances represent the relative intensities of the two lights.

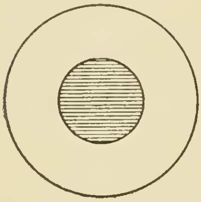

FIG. 413. The Bunsen photometer disc.

CHAPTER LIII.

THE REFLECTION OF LIGHT.

596. The Conditions of Reflection. Whenever a beam of light meets a medium in any way different from the one in which it has been travelling, it is, in general, divided into two parts. One of these parts is turned back into the medium in which it has advanced, while the other is propagated in the second medium, or else is absorbed. That portion of the original beam which is thus turned back and travels again in the first medium is called the reflected beam; the other portion, which enters the second medium, will be considered later. The amount of the original beam which undergoes reflection depends upon the angle at which the light meets the reflecting surface, and also upon the state of polish of this surface. If this polish is as nearly perfect as possible, the reflected light will be confined practically to a single direction, and the mirror itself will appear dark; but if the surface is rough, or if the polish is incomplete, reflections take place in all directions, and we have diffuse or irregular reflection. In general, no surface is so perfectly polished that all of the light is regularly reflected; and, on the other hand, no surface is so rough but that a small portion will undergo regular reflection, if the incident light is nearly parallel to the surface.

597. Explanation of Regular Reflection upon the Undulatory Theory. The wave theory enables us fully to explain the phenomenon of the regular reflection of light, and to deduce the laws according to which this reflection takes place, as may be seen from the following demonstration.

If C is a luminous point, Fig. 414, and $A\ B$ a plane mirror with its reflecting surface upward, we are to inquire as to the position of the wave front of the light after reflection from this mirror occurs. Before reflection, this wave front is a portion of a spherical surface, which has the point C as a center, and which advances towards the mirror AB by Huygens's construction (p. 733). That is, every point upon this wave front becomes a center of disturbance which sends out little spherical waves; and the envelope of all these little secondary waves, which are also known as the Huygens wavelets, becomes the new wave front.

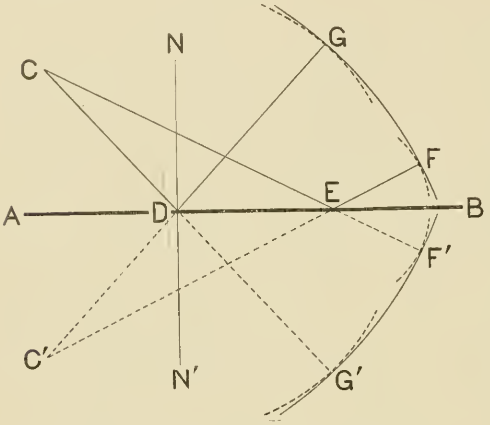

FIG. 414. Explanation of the regular reflection of light upon the undulating theory.

In this way, after the lapse of a certain time, the wave front would have reached the position F' G', if the mirror had not been present. This would have been accomplished by the fact that the disturbances which go out, for instance, from D and E as centers, would then have been found upon the surface of spherical wavelets which had D G' and E F', respectively, as radii. In a similar way, every point between D and E would have contributed wavelets, all of which would then have been tangent to the enveloping wave front F' G'. But since the mirror is present, and, therefore, since the wave propagation cannot take place in the directions D G' and E F', it must remain in the medium above the mirror; and the disturbances which actually go out from D and E as centers are now found upon the circumferences of spherical wavelets which have equivalent distances, namely, D G and E F as radii. Moreover, every point between D and E upon the mirror surface, as the wave front reaches it, becomes a center of disturbance which sends out spherical wavelets, and the envelope of all these secondary wavelets becomes the new wave front F G.

Accordingly, the effect of the reflection from the mirror has been to change the direction of curvature of the wave front, and therefore, also, the direction in which this wave front advances. The wave fronts F G and F' G', one being above the mirror and the other below it, are thus envelopes of spherical wavelets which have equal radii, from which it follows that these wave fronts are *symmetrical* with respect to the surface of the mirror. But $F'G'$ is the portion of a sphere of which C is the center, and therefore $F G$ is the portion of a spherical surface of which C' is the center, where C' is symmetrical with C. The location of the point C' is thus perfectly definite. In other words, C' is behind the reflecting surface, on the perpendicular drawn from C to this surface, and just so far behind this surface as the point C is in front of it. Accordingly, the spherical wave front of the reflected wave travels away from the mirror as if it actually came from the point C'.

598. The Laws of Regular Reflection. If we do not consider the spherical *waves* of light, but the *radii* of these waves, or the *rays* of light, as they depart from C, Fig. 414, it is evident that the ray $C D$, for instance, will be reflected as the ray $D G$, and and the ray $C E$ will be reflected as the ray $E F$, both of these rays, leaving the mirror surface as if they came from C'. The ray departing from the luminous source and striking the mirror is called the *incident ray*, and the ray which leaves the mirror after reflection is called the *reflected ray.* If NN' is drawn normal to the reflecting surface at the point D, *the plane determined by the incident ray and the normal is called the plane of incidence.* Moreover, the angle CDN, formed by the incident ray and the normal, is called *the angle of incidence*, and the angle NDG, formed by the normal and the reflected ray, is called *the angle of reflection.* Inasmuch as the upper part of the figure is in every way symmetrical with the lower, the reflected ray DG must be in the plane of incidence, since the ray DG', which is symmetrical with DG, is in the plane of incidence. We have, therefore, the first law of regular reflection, namely, *the reflected ray remains in the plane of incidence.* Furthermore, since

$$\text{Angle } CDN = \text{Angle } C'DN', \qquad \text{(by symmetry)}$$

and since

$$\text{Angle } NDG = \text{Angle } C'DN' \qquad \text{(vertical, angles)}$$

it follows that

Angle CDN = Angle NDG.

But CDN is the angle of incidence, and NDG is the angle of reflection. We have, therefore, the second great law of regular refletion, namely, *the angle of incidence equals the angle of reflection.*

599. When the Mirror Rotates, the Reflected Ray Moves Through Twice the Angle Through Which the Mirror Turns. Because of the second law of regular reflection, it follows that if the angle of incidence changes by a certain amount, the angle of reflection must change by the same amount. Accordingly, if the mirror is turned through a certain angle, the normal to the mirror will be turned through the same angle, the angle of incidence will then change by this amount, the angle of reflection will also change by the same amount, and the angle between the incident and the reflected ray will change by twice this amount. Therefore, if the incident ray maintains its original direction, the reflected ray moves through an angle twice as great as that through which the mirror turns.

600. Images in Plane Mirrors. We have seen that rays of light, when reflected by a plane mirror, leave this surface as if they came from a point behind the mirror which is symmetrical with the radiant point before it. This point behind the mirror, symmetrical with the luminous point in front, is known as the *image* of the radiant point. The image of a point therefore is defined as the point at which the rays of light meet or seem to meet after reflection. If the rays of light, after leaving the mirror surface, really meet at the image, the image is said to be *real*; but if they do not actually meet at this point, but only appear to do so, the image is said to be *virtual*. The image in a plane mirror is therefore always virtual.

The image of an object is the assemblage of the images of its several points. Since the image of an object is symmetrical with the object, the image is always perfectly definitely located, and in a plane mirror it is always behind the mirror, on the perpendicular from the object to the mirror, and just as far behind the mirror as the object is in front of it. The image, therefore, is just as fixed in position as is the object which produces it, and is wholly independent of the position of the eye by which it is seen. Accordingly, it is a perfectly simple matter to construct images of objects in a plane mirror, since this can be done by constructing the image of several of its points, and connecting these by lines. Thus if MM', Fig. 415, represents a plane mirror, and AB is an object before it, the image of each point of the object, in this case an arrow, may be located by the above rule for the position of images. Since the image of the point A must be symmetrical with the point A itself, it must be behind the mirror, on the perpendicular from the point A to the mirror, and just as far behind the mirror as this point is in front of it. The image A' of the point A, therefore, is fixed in position, and in the same way the image B' of the point B is also fixed. This determines absolutely the location of the image $A' B'$, irrespective of the position of the eye by which this image is seen.

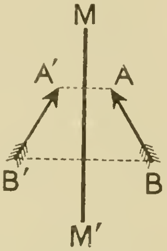

FIG. 415. The construction of an image in a plane mirror.

601. Construction of the Path of the Ray of Light by Which an Image is Seen. Whenever the eye perceives the image of an object, it locates this image, real or virtual, at the point where the rays of light which enter the eye meet or seem to meet. If it is desired to construct the path of the ray of light by which this image is seen, *the image must first be constructed,* and the eye then placed in position. If the eye E, Fig. 416, sees the image of the point of the arrow, the ray of light CE enters the eye as if it really came from the image A', so that we *begin at the eye* and construct the ray of light CE, just as if it really came from the point A', until it meets the surface of the mirror at C. At C the light must undergo reflection; and since A' is the image of the point A, the light which appears to come from A' must really come from A, along the path AC. Of all the rays, therefore, which emanate in all directions from A, that particular ray which passes in the direction AC is reflected at C to the eye at E, so that the path of the particular ray of light by which the

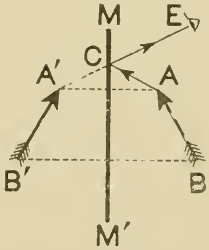

FIG. 416. Construction of the path of the ray of light by which an image is seen.

image A' is seen is ACE. If the eye were in another position, a different ray of light would enter it, and accordingly the path of the ray of light by which the image is seen would then be different. Therefore, while the image is definitely located and fixed, the path of the ray of light which enters the eye depends upon the position of the eye.

602. Multiple Reflections and Multiple Images. When rays of light are reflected from a plane mirror, they depart from that mirror in the direction in which they would proceed if they actually came from the virtual image of the source of light. Thus, if L is a luminous point, and MM a reflecting surface, Fig. 417, we may follow two rays of light from L, for instance LA and LB, which undergo reflection at A and B respectively and advance in the directions AC and BD, precisely as if they came from L' which is the image ~~source of L~~ of the source L.

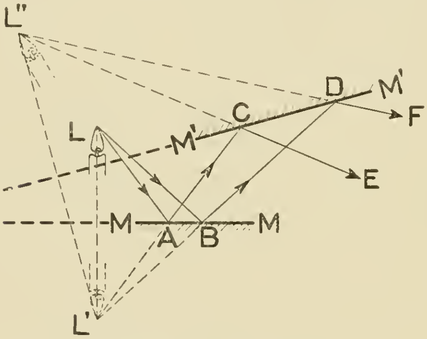

FIG. 417. The principle of multiple reflection and multiple images.

If these rays encounter another mirror $M'M'$, they undergo reflections at this surface also, and leave $M'M'$ in the directions CE and DF. Since the rays AC and BD are incident upon $M'M'$ as if they came from L', the reflections which occur at $M'M'$ take place as if these rays really departed from this image. The point L'', therefore, at which the rays CE and DF apparently meet after reflection, is the image in $M'M'$ of the image L', from which the rays AB and CD apparently come. But L'' is found to be behind the mirror $M'M'$, on the perpendicular from L' upon $M'M'$ and just as far behind $M'M'$ as L' is in front of it. Accordingly L'' may be regarded as the image in the second mirror of the image L' in the first. In this sense an image in one mirror is said to cast an image in another mirror; by which is meant, not that the rays which are reflected at the second mirror actually depart from the image in the first, but that they proceed to the second mirror, and therefore undergo reflection at this mirror, *as if* they came from the virtual image of the original object.

If the rays *CE* and *DF* were to encounter another mirror, it is evident that they would be reflected at that mirror and proceed from it as if they came from the image in that mirror of the image *L″*. *At each reflection, therefore, an image is produced;* so that the phenomenon of multiple reflections, as well as that of multiple images, is occasioned by rays of light which undergo successive reflections from one mirror to another.

The above analysis of multiple reflection shows that the location of the various images takes place in strict accordance with the rule for the location of the image given in Art. 600. In practice, therefore, the application of this rule is found to be sufficient, and the complication introduced by the construction of the rays of light is avoided. When the various images are thus located, the eye may be placed in any position, and the path of the ray of light by which any desired image is seen may easily be traced. Two or three typical instances of the formation of multiple images may be cited.

603. Multiple Images in Two Mirrors Making an Angle of 90°. If two mirrors, with their reflecting surfaces turned towards each other, Fig. 418, make an angle of 90° with each other, an object *A*, placed between them, will, by the former construction, give an image at *B* in the vertical mirror and an image at *C* in the horizontal mirror. Each of these virtual images is the point at which rays of light apparently meet after reflection from the respective surfaces. Thus, if two rays of light, *AG*, are incident upon the horizontal mirror at *G*, they will be reflected in the direction *GF* as if they came from the image *C*. But at *F* these rays strike the vertical mirror, and are here reflected as if they came from *D*, so that *D* is thus seen

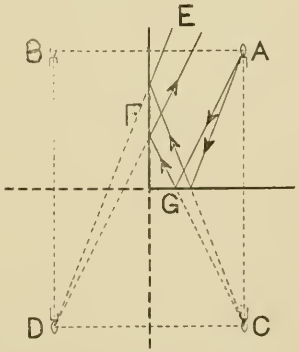

Fig. 418. Multiple images in two mirrors making an angle of 90° with each other.

to be the image in the vertical mirror of the image *C* in the horizontal. In a similar way it might have been shown that *D* is also the image of the horizontal mirror of the image *B* in the vertical.

Without following any particular rays of light, therefore, we might have said, more simply, that just as *B* and *C* are images of *A* in the vertical and the horizontal mirrors respectively, so, also, is *D* the image of *C* in the vertical mirror, and likewise the image of *B* in the horizontal mirror. In other words, *D* is behind the vertical mirror, on the perpendicular from *C* to this mirror, and just as far behind this mirror as *C* is in front of it. Also, *D* is behind the horizontal mirror, on the perpendicular from *B* to this mirror, and just as far behind this mirror as B is in front of it. Moreover, no further reflections are now possible, inasmuch as *D* is behind the reflecting surfaces of both mirrors. An eye placed at *E*, so as to receive the rays *FE*, would perceive the image *D*, where these rays apparently meet.

604. Construction of the Path of the Ray of Light by Which an Image is Seen. When the images are all constructed, the eye may be placed in any desired position, and the path of the ray of light easily followed by which a particular image is seen. Thus, if the eye is placed at *E* and the image is viewed at *D*, Fig. 419, the actual path of the ray of light by which this image is seen may be constructed in the following manner: *Beginning at the eye*, the ray *E F*, is drawn as if it came from the image at *D*, and continued until it strikes the mirror at *F*. At the surface of the mirror at *F* this ray must undergo reflection, and since *D* is the image of *C*, the ray passes from *G* to *F* as if it came from the point *C*. But at *G* the ray must again undergo reflection; and since *C* is the image of the point *A*, the ray which is reflected at *G* must come from *A*. Therefore, the actual path of the ray of light which departs from the object at A and enters the eye

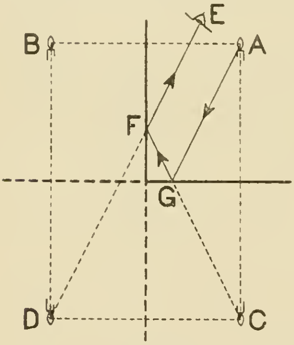

Fig. 419. Construction of the ray of light by which a particular image is seen.

at *E*, thus enabling the image at *D* to be seen, is from *A* to *G* *G* to *F*, *F* to the eye at *E*.

605. Multiple Images in Mirrors Making an Angle of 60°. Let the mirrors *M* and *M′*, Fig. 420, be inclined to each other at an angle of 60°, and let an object *A* be placed in the opening between them. By applying the principles for the location of images deduced in Art. 602, we see that the object *A* gives an image at *B* in the mirror *M* and also an image at *C* in the mirror *M′*. The image *B* in the mirror *M* also casts an image *D* in the mirror *M′*, just as if *B* were a real object. Each of the images *C* and *D*, being before the the mirror *M*, casts an image in this mirror, forming the images *E* and *F*. The image *F* being now before the mirror *M′* casts an image in that mirror which coincides with the image *E*, if the mirrors are at an angle of 60°. Since *E* is now behind both mirrors, no further reflections can take place. Moreover, it is easy to show that these images are all upon the circumference of a circle of which the point of intersection of the mirrors is the center.

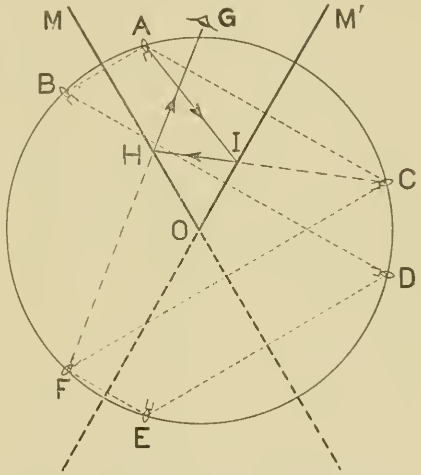

FIG. 420. Multiple images in two mirrors making an angle of 60° with each other.

When the images have all been located, the eye may be placed in the opening between the two mirrors, and the path of the ray of light whereby any particular image is seen may be easily traced. If, for instance, the eye is placed at *G*, and the image is observed at *F*, *we begin at the eye* and trace the ray *GH*, as if it actually came from *F*. At *H*, however, this ray undergoes reflection; and as *F* is the image of the image *C*, the ray advances from *I* to *H*, as if it came from *C*. But at *I* the ray is again reflected; and as *C* is the image of the object *A*, the ray comes to *I* directly from *A*. Therefore, the ray of light which departs from *A* and travels in the directions from *A* to *I*, from *I* to *H*, and from *H* to *G*, is the particular ray by which the eye at *G* perceives the image at *F*.

The number of the images, produced in two mirrors, making an angle of 60° with each other, is thus seen to be six, including the original object. In general, the number of images which can be formed in two mirrors making an angle with each other is the entire circumference, or 360°, divided by the angle which the mirrors make with each other. The number of images thus given, however, includes the original object.

The action of kaleidoscope depends upon the principle of multiple reflection in two mirrors making an angle with each other.

606. Multiple Images in Parallel Mirrors. When an object is placed between two parallel mirrors, the angle between these mirrors is zero, and the number of images is theoretically infinite. Since an image is always upon a perpendicular from the object to the mirror, these images are located upon a line which passes through the original object perpendicular to the planes of the mirrors, this line being indefinite in extent in both directions. The object *A*, Fig. 421, gives an image *B* in the mirror *M* and an image *C* in the mirror *N*. The image *B* likewise gives the image *D* in the lower mirror. *C* and *D* are now reflected in the upper mirror, giving the images *E* and *F* respectively, and these in turn give the images *G* and *H* in the mirror *N*. In the

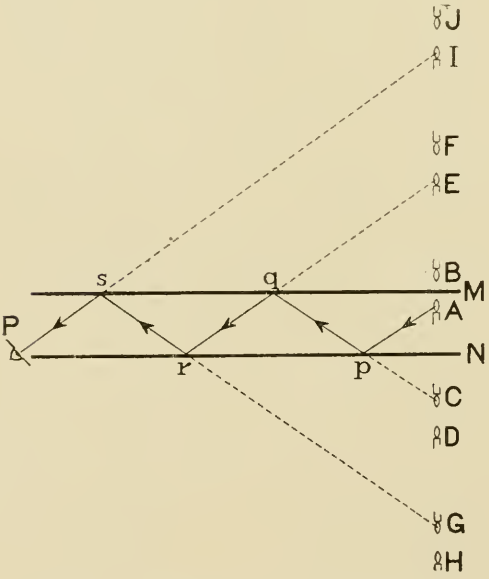

FIG. 421. Multiple images in parallel mirrors.

same way *G* and *H* give the images *I* and *J* in the mirror *M*, and so on indefinitely.

To construct the path of the ray of light by which any image, as, for instance, *I* is seen by an eye placed at the point *P*, we begin at the eye and construct the ray *sP* as if it came from the image *I*. At *s* the ray undergoes reflection, and since *I* is the image of *G*, the light comes from *s* along the line *rs*, as if it really came from *G*. But *G* is the image of *E*, hence the light comes to *r* along *qr*, as if it came from *E*. *E* being the image of *C*, light comes to *q* along the line *pq*, as if it came from *C*; and since *C* is the image of the object *A*, the light which is reflected at *p* comes to this point along the line *Ap*. It follows, therefore, that the path of the particular ray of light which enters the eye at *P*, by which the image *I* is seen, is *ApqrsP*.

607. Spherical Mirrors. A spherical mirror is a portion of a sphere, so silvered as to produce a reflecting surface. The mirror thus obtained is concave if the interior of the spherical surface is silvered, and convex if the exterior portion of the sphere is the reflecting surface. The center of the sphere of which the mirror forms a part is called the *center of curvature* of the mirror. The *aperture* of the mirror is the solid angle subtended at the center of curvature by the reflecting portion of the sphere. The central portion of the spherical reflecting surface is known as the *vertex* of the mirror.

Any line drawn through the center of curvature and meeting the mirror is called an *axis* of the mirror. Since an axis is always the radius of the sphere of which the mirror forms a part, an axis is always perpendicular to the surface of the mirror. *That particular axis passing through the vertex of the mirror is called the principal axis, while all others are secondary axes.*

The focus of a spherical mirror is the point at which the rays of light meet after reflection. If the source of light is at an infinite distance, so that the rays which fall upon the mirror come *parallel to the principal axis*, the focus at which they meet after reflection is known as the *principal focus*. *Conjugate foci are two points so related that rays of light departing from one will be brought to a focus at the other.*

608. Location of the Focus in a Spherical Mirror. The conditions which determine the location of a focus with respect to a mirror may be determined from the following consideration. Let *mn*, Fig 422, be a portion of a spherical reflecting surface, the center of curvature of which is *C*, and let *L* be a radiant point. If we imagine the light from *L* to be divided into separate rays, *LB* will be one of these rays meeting the mirror. Since the radius *CB* is normal to the reflecting surface at *B*, the ray of light which strikes the mirror at *B* will be reflected to a point *l* on the principal axis, so that the angle of incidence *LBC*, equals the angle of reflection *lBC*. In the triangle *LBl*, therefore, we have the line *BC* bisecting the angle at the vertex *LBl*. From the proposition in geometry, which states that the line bisecting the angle at the vertex of a triangle divides the base into segments proportional to the adjacent sides, we have

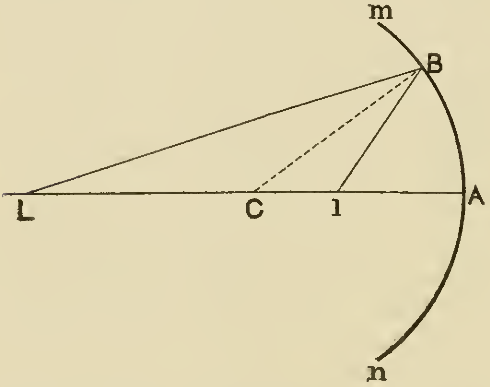

Fig. 422. Location of the focus in a concave mirror.

$$\frac{LC}{lC} = \frac{LB}{lB}.$$

If we now have to deal with a mirror of small aperture, that is, if we consider but a very small area of the spherical surface, we may take a portion so small that *LB* practically equals *LA*, and *lB* practically equals *lA*. If we make these approximations, as we are justified in doing when the aperture is small, the above equation becomes

$$\frac{LC}{lC} = \frac{LA}{lA},$$

or

$$\frac{LC}{LA} = \frac{lC}{lA} \cdots\cdots\cdots\cdots\cdots\cdots (A)$$

But the ratio

$$\frac{LC}{LA}$$

is constant, and hence

$$\frac{lC}{lA}$$

must also be constant. Therefore, the point l must be constant, and l is fixed in position when L is fixed. Accordingly, l is the focus corresponding to the radiant L.

609. The Explanation of Foci. To apply the undulatory theory to the explanation of foci in spherical mirrors, we must imagine the wave front *ab*, Fig. 423, which departs from the source of light L, as it would have existed if the mirror *mn* had not been present. As the wave front reaches the mirror, every particle of the surface of the mirror becomes a center of disturbance sending out little spherical waves. With every point of the mirror, therefore, as a center, we describe little spherical wavelets tangent to the wave front *ab*, and the envelope of all these elementary wavelets will give the new wave front *a′ b′*. This new wave front

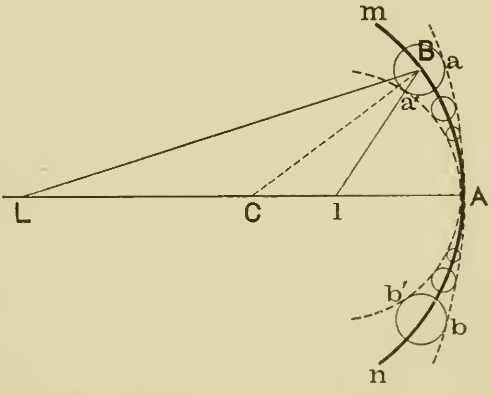

FIG. 423. Reflection from a spherical mirror explained by the undulatory theory.

is on the opposite side of the mirror, and since the aperture of the mirror is small, it does not differ widely from the portion of a sphere having its center at *l*. The reflected wave, therefore, converges to a point at *l*, from which it follows that this point is the focus corresponding to the luminous point *L*. The focus, therefore, is a point to which the spherical wave converges, or shrinks, after reflection. If the radiant were placed at *l*, *lB* would then be the incident ray and *BL* the reflected ray, and the wave which departs from *l* would converge at *L*, so that *L* is the focus corresponding to the point *l*. *L* and *l* are therefore at conjugate foci.

610. Relation Between the Conjugate Foci of a Spherical Mirror. In determining the relation which exists between the positions of the conjugate foci with respect to the mirror, we shall assume that all distances to the left of the mirror, that is, on the side of the mirror towards the source of light are positive, Fig. 422, while all distances to the right, or on the side away from the source of light, are negative.

In equation (A), letting

$$LA = p,$$

$$lA = p',$$

$$CA = r,$$

we have

$$\frac{p - r}{p} = \frac{r - p'}{p'},$$

where *r* is the radius of the curvature. From this,

$$pp' - p'r = pr - pp',$$

or

$$p'r + pr = 2\,pp'.$$

Dividing by $pp'r$,

$$\frac{1}{p} + \frac{1}{p'} = \frac{2}{r}.$$

This expression shows the relation between the conjugate foci, and is called the fundamental equation of the mirror. A number of special cases are important, depending upon the distance of the luminous point from the reflecting surface.

1. When the radiant point is at infinity, that is, when the rays come parallel to the principal axis. In this case

$$p = \infty,$$

and p' is the distance of the principal focus from the mirror. The fundamental equation of the mirror then becomes

$$\frac{1}{\infty} + \frac{1}{p'} = \frac{2}{r},$$

or

$$\frac{1}{p} = \frac{2}{r},$$

whence

$$p' = \frac{r}{2}.$$

Therefore, the principal focus of a concave mirror is situated half-way between the center of the curvature and the mirror.

2. When the radiant point is at the center of the curvature, that is, when

$$p = r.$$

Then, from the fundamental equation,

$$\frac{1}{r} + \frac{1}{p'} = \frac{2}{r},$$

or

$$\frac{1}{p'} = \frac{1}{r},$$

and

$$p' = r.$$

Therefore, if rays depart from the center of curvature, they return to this point after reflection.

3. When the radiant point is at the principal focus, that is, when

$$p = \frac{r}{2}.$$

Then, from the fundamental equation,

$$\frac{2}{r} + \frac{1}{p'} = \frac{2}{r},$$

or

$$\frac{1}{p'} = 0,$$

and

$$p' = \infty.$$

Therefore, when the radiant point is at the principal focus, the reflected rays leave the mirror only to meet at infinity, that is, they leave the mirror parallel to the principal axis.

4. When the radiant point is between the principal focus and the vertex of the mirror. In this case

$$p < \frac{r}{2},$$

and

$$\frac{1}{p} > \frac{2}{r}.$$

The equation for the mirror then remains

$$\frac{1}{p} + \frac{1}{p'} = \frac{2}{r},$$

or

$$\frac{1}{p'} = \frac{2}{r} - \frac{1}{p}.$$

But, since $\frac{1}{p} > \frac{2}{r}$, it follows that $\frac{1}{p'}$ is negative, and the focus corresponding to this position of p is on the other side of the mirror, and is virtual. The rays, therefore do not converge after reflection, but continue to diverge, as if they came from the virtual focus behind the mirror.

Whatever positions p and p' occupy, they are always connected by the fundamental equation of the mirror, and are always at conjugate foci. A construction similar to the one here given for the principal axis applies also to any secondary axis.

In the case of convex mirrors, r is negative, and therefore p' is also negative. The focus of a convex mirror is therefore always virtual.

611. Spherical Aberration—The Astigmatism of the Concave Mirror. When the aperature of a spherical mirror is small, we have seen that rays of light which depart from a source L on the principal axis, Fig. 422, are brought to a focus, l, likewise on the principal axis. The smaller the aperture of the mirror, the more sharply defined is the focus l. When, however, two rays, such as LA and LB, fall obliquely upon a mirror of large aperture, Fig. 424, they cross each other at a point l, which is not upon the principal axis, but which is the farther removed from this line the greater the angle of incidence. After crossing at l_1, these rays then intersect the principal axis at l_2, since this axis is always in the plane of incidence.

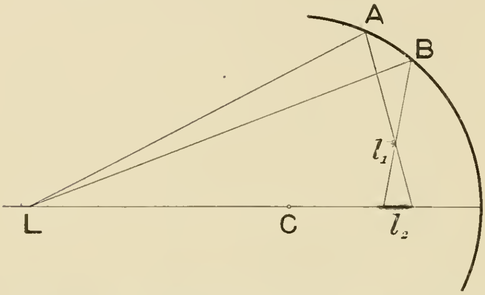

FIG. 424. Primary and secondary focal lines, formed by reflection from a concave mirror of large aperture.

Moreover, if we consider a portion of the mirror similar to AB but slightly towards the observer from the plane of the paper, that is, if we think of the figure as rotated through a small angle about LC as an axis, AB will describe an element of the mirror surface, the point l_1 will trace a short line, which is really the arc of a circle, while l_2 will remain in the same position. Accordingly the light from a point at L, as it falls obliquely upon a portion of the mirror at AB, is brought to a focus *not at a point, but along two lines*, one l_1, which is known as the *primary focal line*, and the other, l_2, which is known as the *secondary focal line*. At some point between l_1 and l_2, the cone of light reflected from the mirror surface must pass through an area, approximately of equal width in each direction, known as the *circle of least confusion*. This is the nearest approach to a single focus which the concave mirror is capable of producing.

This failure of a spherical mirror of large aperture to bring to a sharp focus light which falls upon it is called the *spherical aberration* of the concave mirror. The meeting of the reflected rays along two lines at right angles to each other, even when the source of light is merely a point, necessarily gives rise to an indistinctness, or blurring, of the image, and produces the phenomenon known as astigmatism.

If we consider the wave front of the reflected light when astigmatism occurs—this wave front being then called an *astigmatic wave front*—we shall find that this is not spherical, but is more sharply curved in one direction than in a direction at right angles to this. Thus, if the shaded portion of Fig. 425 represents an astigmatic wave front advancing towards us and to the right, its curvature about the horizontal axis is greater than that about the vertical. This becomes immediately obvious from the fact that the rays from the top and bottom of the wave front are brought to a focus at the primary focal line, l_1, before the rays from the sides unite to form a focus at the secondary focal line, l_2. The entire portion of the figure between l_1 and l_2 may be regarded as the focus, the circle of least confusion being indicated in this diagram by the particular area having a equilateral cross-section.

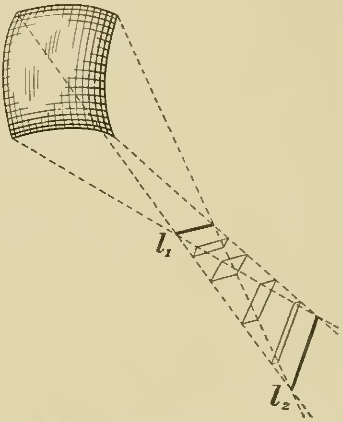

FIG. 425. An astigmatic wave front, converging towards the primary and secondary focal lines.

612. Caustics by Reflection. By noting the various points at which *consecutive* rays cross each other, as they are reflected from the entire surface of the mirror, Figs. 426 and 427, we

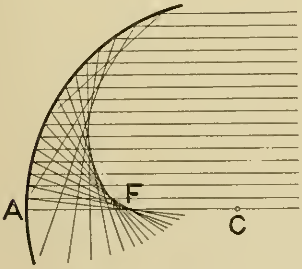

FIG. 426. The formation of the caustic.

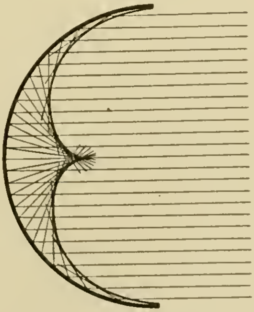

FIG. 427. The caustic by reflection.

shall find that these points all lie upon a curve which is known as the *caustic by reflection*. This curve has a *cusp* at *F*, where *F* is the focus corresponding to the source *L* when the aperture of the mirror is very small. If the figure is rotated about the principal axis, this caustic will generate a surface of complex form, all of which is the focus conjugate with the point-source *L*.

613. Elliptical and Parabolic Mirrors. On account of the spherical aberration and the astigmatism of spherical mirrors, these mirrors cannot be employed to give sharp images. There are, however, two forms of reflecting surfaces which can be used for bringing rays of light absolutely to a single focus, namely, mirrors having an elliptical form, and those having a parabolic form. From a property of the ellipse, rays of light emanating from one focus, Fig. 428, are reflected absolutely to a point at the other focus. Moreover, from a property of the parabola, rays of light parallel to the principal axis are brought absolutely to a point at the focus, Fig. 429. Conversely, rays of light which emanate from a source placed at the focus of a parabolic mirror depart from the mirror in a beam which is parallel to its principal axis.

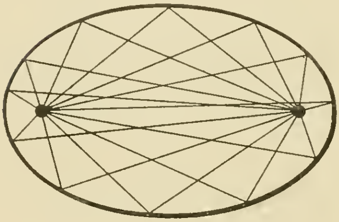

FIG. 428. The principle of the elliptical mirror.

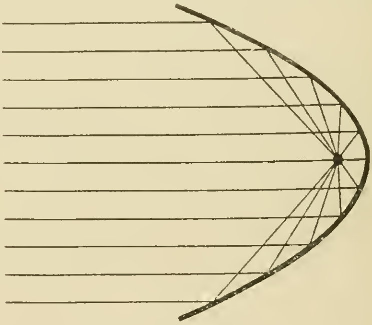

FIG. 429. The parabolic mirror.

614. The Construction of Images in Spherical Mirrors. We have seen that the *image* of a luminous source is the point where the rays of light departing from that source meet after reflection, from which it follows, in the foregoing analysis, that the point *l*, Fig. 422, is *the focus* and at the same time *the image* of the radiant point *L*. Moreover, the image of an object is the assemblage of the images of its several points, and hence, if an extended object were placed at *L*, the image of this object would be found at *l*. But the path of a ray of light is always reversible, that is, it can be made always to retrace its path, so that if *l* were the luminous source, sending out rays of light, these rays of light would be brought to a focus at *L*. *L* and *l* are therefore at conjugate foci; and, in general, *an object and its image are always at conjugate foci.* The relative movements of these conjugate foci, and the size as well as the location of the images cast by a spherical mirror, may best be studied by actually constructing images in this mirror when the object occupies successively a number of positions.

In this construction, it must be borne in mind that the center of curvature is the center of the sphere of which the mirror forms a part, and that any line passing through the center of curvature intersects the surface of the mirror normally. Any ray of light, therefore, passing through the center of curvature strikes the mirror at perpendicular incidence and is reflected back upon itself. It must also be borne in mind that the principal focus is that point upon the principal axis where incident rays which are parallel to the principal axis meet after reflection. With these two facts in mind, together with the fixed location of these points with reference to the mirror, the construction of images becomes at once simple and definite. A number of special cases must be considered depending upon the location of the object with relation to the mirror.

1. When the object is outside the center of curvature, that is, farther from the mirror than the center of curvature. Let the object *XY* Fig. 430, be placed before the concave mirror *mn*, *C* being the center of curvature, and *F* the principal focus. Of the infinite number of rays which depart from the point *X*, one ray, and only one ray, namely *XD*, passes parallel to the principal axis, and this is reflected through the principal focus. Therefore, somewhere on the line *DF* is found the image of the point *X*. So also, of all the rays which emanate from *X*, one ray, and only one ray, namely *XC*, passes through the center of curvature, and this, after striking the mirror perpendicularly at *G*, is reflected

back upon itself. Therefore, somewhere on the line GX is found the image of the point X. Inasmuch, then, as this image is found both on the line DF and also on the line GX, it must be located at their point of intersection x. Therefore, the rays of light which depart from X meet at x, and x, accordingly, is the image of the point X.

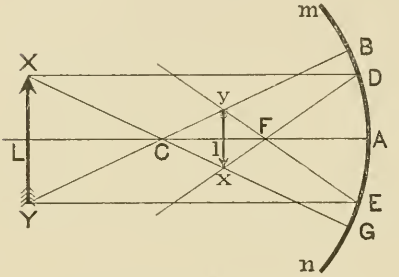

Fig. 430. Construction of an image in a concave mirror when the object is beyond the center of curvature.

The same construction follows for the other end of the object. One ray, and only one ray, from the point Y falls parallel to the principal axis, namely, the ray YE, and this is reflected through the principal focus. Therefore, the image of the point Y is found somewhere on the line EF. Moreover, one ray, and only one ray, from the point Y passes through the center of curvature, namely, the ray YC, and this is reflected at B directly back upon itself, since it strikes the mirror perpendicularly. Therefore, the image of the point Y is found somewhere on the line BC. Accordingly, since the image of Y is found on the line EF and also on the line BC, it must be found at their point of intersection y. Therefore, xy is the image of the object XY, and is located between the center of curvature and the principal focus. Moreover, the image is *real,* since the rays really meet after reflection, it is inverted, and is smaller than the object.

To determine the relative sizes of the object and image, which, as we have seen above, are located at the conjugate foci L and l respectively, we may proceed as follows: From similar triangles,

$$\frac{XY}{xy} = \frac{LC}{lC}.$$

But we have seen that

$$\frac{LC}{lC} = \frac{LA}{lA},$$

and therefore

$$\frac{XY}{xy} = \frac{LA}{lA}$$

Accordingly, the size of the object is to the size of the image as the distance of the object from the mirror is to the distance of the image from the mirror. This law always holds for the relative sizes of object and image in a spherical mirror.

2. When the object is at the center of curvature. In this instance it is necessary to adopt a slightly different procedure, as follows: Of the entire number of rays which emanate from the point X of the object, Fig. 431, the particular ray which passes parallel to the principal axis, namely, XD, is reflected through the principal focus, so that the image of X is found somewhere on the line DF. Also the particular ray XF, which passes through the principal focus, is reflected at E parallel to the principal axis. The point x at which these two rays meet after reflection, is the image of X. Similarly, the point y is located as the image of Y. Accordingly, when an object beyond the center of curvature moves towards the mirror, the image advances to meet it; and when the object is at the center of curvature, the image is likewise at that point, is real, is inverted, and, since image and object are then equally distant from the mirror, is of the same size as the object.

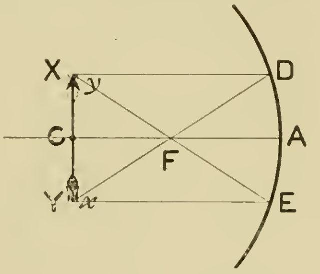

FIG. 431. When the object is at the center of curvature, the image is also at the center of curvature.

3. When the object is between the center of curvature and the principal focus, the image is outside the center of curvature. In this case, the construction of the image is given in Fig. 432, in which rays corresponding to those in Fig. 430, have received the same lettering as in that figure, and accordingly, the same wording in the demonstration will apply equally well to this case. Therefore, when the object is between the center of curvature and

the principal focus, the image is outside the center of curvature, is real, is inverted, and is larger than the object.

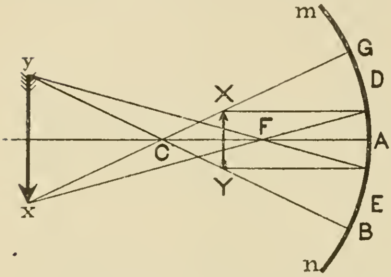

Fig. 432. When the object is between the center of curvature and the principal focus, the image is outside the center of curvature.

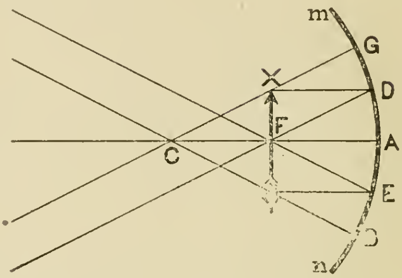

Fig. 433. When the object is at the principal focus, the image is at infinity.

4. When the object is at the principal focus the image is at infinity, and the rays, which in other instances meet to form the image, depart from the mirror parallel to each other, as shown in Fig. 433.

5. When the object is between the principal focus and the mirror, the same construction and the same wording hold as in the other instances. That is, Fig. 434, the ray *XD,* parallel to the principal axis, is reflected through the principal focus, and therefore somewhere on the line *FD,* or *FD* produced, is found the image of the point *X.* Moreover the ray *XG* strikes the mirror normally, and so is reflected through the center of curvature. Therefore, somewhere on the line *CG,* or *CG* produced, we shall find the image of the point *X.* But the lines *CG* and *DF* diverge after reflection, and therefore they seem to meet at the point *x*, which is the virtual image of the point *X.* A similar construction follows that for the point *Y,* locating the virtual image at *y.* Accordingly, *xy* is the image of the object *XY;* it is behind the mirror, is virtual, erect, and magnified.

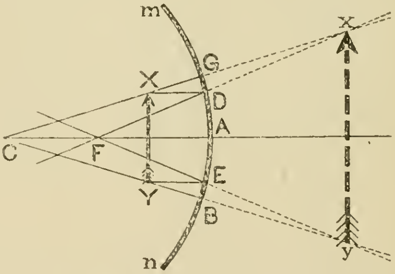

Fig. 434. When the image is between the principal focus and the mirror, the image is behind the mirror, is virtual, erect, and magnified.

6. When the object now approaches the mirror, the image also approaches the mirror, and when the object is at the mirror surface, the image is also at the mirror surface, and coincides with the object. The image is then erect, virtual, and of the same size as the object.

7. When the object passes to the other side of the mirror, it is then outside the sphere of which the mirror forms a part, and we have reflection from a convex surface. In this case, precisely the same construction is to be followed as in the previous cases. If *XY*, Fig 435, is the object, the ray *XD*, passing from the point *X* parallel to the principal axis, is reflected as if it came from the principal focus. Therefore, somewhere on the line *FD* we shall find the image of the point *X*. So, also, the ray *XG* which strikes the mirror normally, that is, as if it were to pass through the center of curvature, is reflected back upon itself, and somewhere on the line *GX*, or on this line produced, we shall find the image of the point *X*. Since this image is found on the line *DF*, and also on the line *CG*, it must be found at their point of intersection *x*, and *x* is the virtual image of the point *X*. In a similar manner, *y* is found to be the virtual image of the point *Y*. When an object is before a convex mirror, the image is thus seen to be on the other side of the mirror, erect, virtual, and smaller than the object.

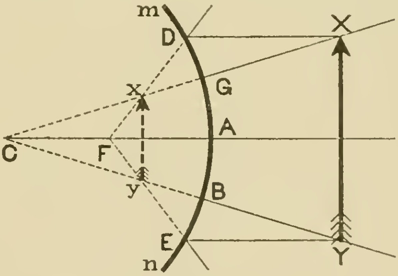

Fig. 435. The image in a convex mirror is always virtual.

In all of these constructions it is to be observed that whenever the image is real, it is always inverted, and whenever it is virtual, it is always erect. This is found to be true of all images produced by whatever means, and therefore *a real image is always inverted, and a virtual image is always erect.*

That the rays of light really do meet before a concave mirror, giving a real, inverted image of the object, may be beautifully shown by placing an inverted bouquet of flowers at the center of curvature of a concave mirror, and tipping the mirror so that the real, inverted image of the bouquet also appears at this same point. The bouquet appears, therefore, erect, and perfectly natural, producing the optical illusion known as the "phantom bouquet."

615. The Japanese Mirror. The Japanese mirror produces a most interesting phenomenon in the reflection of light. This mirror is a polished plate of metal, with a figure, or design, cast in relief upon the back, thus causing the mirror to be thicker in some places than in others. When the mirror is polished, the thin places bend and yield slightly under the polishing, so that the surface, when finished, is not a true plane, but is made up of convexities and concavities, which, indeed, are so slight as not to be noticeable when the mirror is held in the hand. When, however, a beam of sunlight is reflected from this mirror, the rays of light receive a slight inclination in one direction from some parts of this uneven surface, and in the opposite direction from others, a convergence of the rays being produced by the thicker portions of the mirror and a divergence occurring from those portions where the metal yielded under the polishing, that is, where the mirror is thinner. These converging and diverging rays of light give rise to areas of unequal illumination upon the screen; and as this inclination of the rays is ultimately occasioned by the varying thickness of the metal, it follows that, in the beam of light reflected upon the screen, the pattern which exists in relief upon the back of the mirror becomes at once visible. If the mirror is heated, it expands unequally because of the differences in its thickness, and the figure becomes more distinct. The figure, or pattern, may also be made to appear more clearly if the mirror is made the front surface of a reservoir into which air can be pumped. The pressure of the air upon the mirror causes a bending of the thinner portions, and the figure appears more sharply defined.

It is possible, however, that the phenomenon of the Japanese mirror is more complicated than the one here set forth. From the fact that the merest scratch upon the back becomes at once visible in the light reflected from the surface, it would appear that the cast surface of the metal is under a tense strain, and that this strain is relieved when the surface is in any way broken. The yielding of the mirror when this strain is thus changed produces corresponding convexities and concavities in the reflecting surface, and a scratch upon the back or upon the figure which is there in relief, becomes apparent in the reflected light.

The reflection occurring in cylindrical and conical mirrors are very complicated, and the distortions of the images which are thus produced are often grotesque.

CHAPTER LIV.

THE REFRACTION OF LIGHT.

616. Nature of Refraction. We have seen that whenever light meets a transparent medium in any way different from the one in which it has been travelling, it is in general divided into two parts. One portion of the original ray is turned back into the medium in which it has advanced, forming the *reflected* ray, while the other part is propagated in the second medium, and is known as the *refracted* ray. Thus, if *mn*, Fig. 436, represents the boundary between air and water, a ray of light *AB* in the air will, in general, be divided into two parts as it strikes the water, that is, into the reflected ray *BC*, which continues in the air, and into the refracted ray *BD*, which is propagated in the water. It is with the latter, or the refracted ray, that we now have to deal.

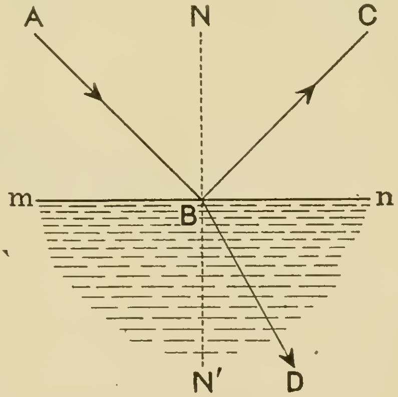

FIG. 436. The reflected and the refracted rays.

In this connection it is to be noted that the refracted ray *BD* does not continue in the same direction as the incident ray, but that it undergoes a bending at the point B where it enters the second medium. The fact of this bending is the origin of the term *refraction*, which is the name given to this phenomenon. *Refraction is the bending which a ray of light undergoes when it passes obliquely from one medium to another.* Moreover, when the light passes from an optically rarer to an optically denser medium, as in the present instance when it passes from the air into water, the refraction, or bending, takes place towards the normal NN'. The angle which the refracted ray makes with the normal, that is the angle DBN', is called the angle of refraction.

617. Light Travels More Slowly in a Denser Medium. Before investigating the relation which the angle of refraction bears to the angle of incidence, it is all-important to ascertain the velocity with which light is propagated in a denser medium, such as water, as compared with its velocity in the air. This determination was made experimentally by Foucault as an adjunct to his celebrated experiment of the measurement of the velocity of light in air (p. 740). Between the rotating mirror M, Fig. 437, and the fixed mirror N of his apparatus, he placed a long tube filled with water, so that the light, after leaving the rotating mirror M, passed twice through the column of water before returning to this mirror. By actual measurement, Foucault then found that light travels more slowly in water than in air, and that its velocity in the former medium is only about $\frac{3}{4}$ as great as in the latter. Even if the phenomenon of interference, as discovered by Thomas Young, had not firmly established the wave theory of light, the experimental determination by Foucault that light moves more slowly in wa-

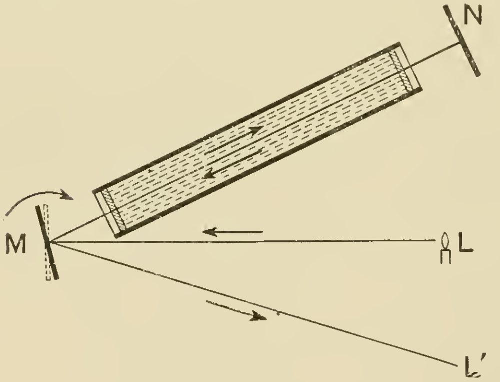

FIG. 437. Foucault's measurement of the velocity of light in water.

ter than in air would forever have done away with the corpuscular theory, since the arguments in support of the latter hypothesis demanded that light should move faster in the denser medium.

618. The Corpuscular Theory Fails to Explain the Retardation in the Denser Medium. It was formerly held that, as a corpuscle, moving from *A* to *B* with the velocity of light, approached the surface of the water *mn*, Fig. 438, the water would attract it, and that this attraction would act downward, that is, normal to the surface. This vertical force of attraction *BF* could then be resolved into two components, one, *BE*, in the direction of motion, and the other, *BN*, at right angles to this direction. The component *BN* would deviate the corpuscle from its original path and make it bend towards the normal in the direction *BD*, while the component *BE*, acting in the direction of motion, would accelerate it, and thus make it move faster in the water than in the air. The first of these results, that is, the bending of the ray of light towards the normal in the denser medium, actually does take place, while the second, namely, the supposition that light must move faster in the denser medium, is not confirmed by Foucault's experiment. Accordingly, Foucault's discovery that light moves not faster, but slower, in the denser medium proves that the attraction theory cannot be true, and therefore that the corpuscular theory cannot account for the phenomenon of refraction.

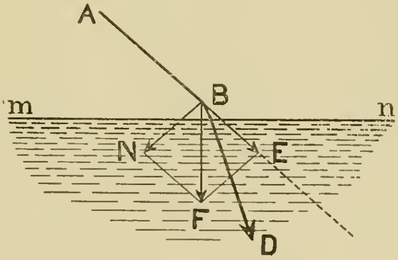

FIG. 438. The refraction of light upon the corpuscular theory.

619. Refraction Explained Upon the Wave Theory—The Index of Refraction. We have, accordingly, to explain the phenomenon of refraction upon the wave theory, and this can be done with the aid of Huygens's construction, as follows:

Let the upper part of Fig. 439 represent air and the lower part water, *mn* being the boundary between these two media. Let us assume that *CC′C″* represents a beam of parallel light, and over, let *AB* be the position which the wave front would have occupied had the wave continued in air. But at *AG* the beam meets the surface of the water, and in this denser medium the light can travel only $\frac{3}{4}$ as fast as in the air. Accordingly, the disturbance which starts out from *G*, and which would have been found upon the surface of a sphere of which *GB* is the radius, is now found upon the surface of a sphere of which *GF* is the radius, where *GF* is only $\frac{3}{4}$ *GB*. In like manner, each point in the refracting surface *GA* must be considered as the center of two systems of spherical waves, one of these systems being everywhere tangent to *AB*, and the other being drawn with radii only $\frac{3}{4}$ as great. The envelope of this second system of spheres is the new wave front *AF*.

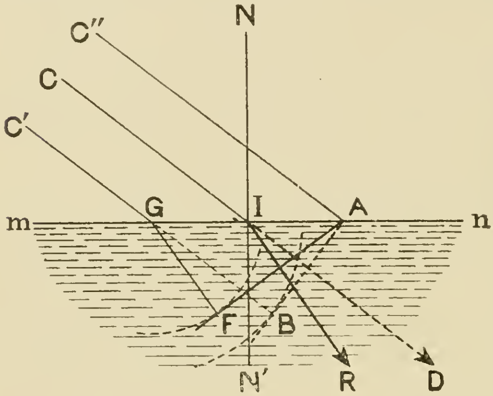

FIG. 439. Refraction of light explained upon the wave theory.

The incident ray of light *CI* is therefore not propagated in the direction *ID*, perpendicular to the wave front *AB*, as would have been the case if the water had not been present, but it advances in the direction *IR*, perpendicular to the wave front *AF*, as this exists in the denser medium. *CIN*, therefore, is the angle of incidence, and *RIN′* is the angle of refraction, where *NN′* is normal to the water surface at *I*.

Since the light would have travelled the distance *GB* in the air in the time it has actually travelled the distance *GF* in water, the lengths of these lines are proportional to the velocities of light in the air and in water respectively. Accordingly,

$$\frac{GB}{GF} = \frac{\text{velocity of light in air}}{\text{velocity of light in water}}.$$

This ratio, however, is a constant quantity, usually denoted by n, and is known as the index of refraction. Therefore, *the index* *of refraction of any medium with reference to air is defined as the ratio of the velocity of light in air to the velocity of light in that medium.* The index of refraction of any medium with reference to the pure ether may be obtained by multiplying its index with reference to air by 1.000292, which is the ratio of the velocity of light *in vacuo* to that in air.

From the relation which the angles of incidence and refraction bear to each other, the numerical value of the index of refraction may be obtained as follows: In the triangle GBA, the angle GAB is equal to the angle of incidence i, since its sides are perpendicular respectively to the sides of the angle of incidence CIN; and in the triangle GFA, the angle GAF is equal to the angle of refraction r, since its sides are perpendicular respectively to the sides of the angle of refraction RIN'. Therefore,

$$\frac{GB}{GA} = \sin i,$$

and

$$\frac{GF}{GA} = \sin r.$$

Dividing the first equation by the second,

$$\frac{GB}{GF} = \frac{\sin i}{\sin r}.$$

But, as we have seen above,

$$\frac{GB}{GF} = \frac{\text{velocity of light in air}}{\text{velocity of light in water}} = \text{index of refraction.}$$

Accordingly,

$$\text{the index of refraction, or } n = \frac{\sin i}{\sin r}.$$

In general, when light passes from a rarer to a denser medium, and when the angle of incidence is made to vary, the angle of refraction will likewise vary, but always in such a way that

$$\frac{\sin i}{\sin r} = n,$$

where n is a constant quantity for these two media. This expression is known as Snell's, or, sometimes, Descartes's, law of refraction.

What has been said regarding the refraction which occurs when light passes from air into water is true whenever light enters a medium optically denser than the one in which it has been travelling. The velocity of light is less in the denser medium than in the air, consequently the index of refraction is greater than unity, and the refracted ray is bent towards the normal. Since, however, the passage of a ray of light is reversible, that is, since light may be made to retrace its path in the opposite direction, it is evident that if the ray of light is sent from the denser to the rarer medium in the direction *RI,* the light is refracted at the boundary between the two media, passing out into the rarer medium in the direction *IC.* Therefore, *when light passes from a denser to a rarer medium, the refracted ray is bent away from the normal.* In this instance the angle *RIN′* becomes the angle of incidence *i,* and the angle *CIN* the angle of refraction *r.* Accordingly, for refraction from a denser to a rarer medium

$$\frac{\sin i}{\sin r} = \frac{1}{n},$$

where n is the index of refraction from the rarer to the denser medium. For refraction from the denser to the rarer medium, the index of refraction is less than unity, and the refracted ray is bent away from the normal.

620. Illustrations of Refraction. Illustrations of refraction abound. If a coin is placed in the bottom of a vessel, Fig. 440, and the eye is brought to such a position that it is invisible, it is apparently lifted up and may readily be seen when the vessel is filled with water. Rays of light departing from the coin undergo refraction at the surface of the water, and are bent away from the normal. Two rays *AB* and *AC,* therefore, will be so bent at the surface that they can enter the eye at *E*. The elevated image *D*, at which these refracted rays of light seem to meet, is the point to which the eye refers the location of the coin.

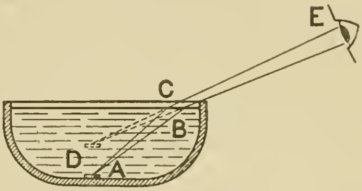

Fig. 440. A coin in a vessel of water becomes visible through the refraction of light.

In the same way a stick appears bent when thrust obliquely into water. Rays of light emanating from the end of the stick, Fig. 441, undergo refraction at the surface and are bent away from the normal. Two of these rays, departing from *A* in the direction *AB* and *AC*, are refracted so as to enter the eye at *E*. The eye, however, always refers the image of an object to the point of meeting of the rays of light which enter the eye; and therefore, since these rays converge at the point *D*, the end of the stick seems elevated above its true position and appears at*D*. In the same way, each point of the stick beneath the surface of the water is apparently lifted up, and hence the stick seems to be broken at the point *G* where it enters the water.

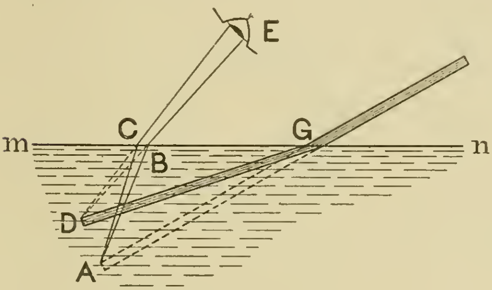

Fig. 441. A stick thrust into water appears broken, due to the refraction of light.

In virtue of our binocular vision, a vessel filled with water appears to be of less depth than it really is, even when we are looking vertically into the water. If *E* and *E′*, Fig. 442, represent the eyes of the observer, rays of light from a point *A* at the bottom of the vessel pass in the direction *AB* and *AC*, and also in the direction *AB′* and *AC′*. At the surface they undergo refraction, being bent away from the normal, and thus enter the eyes at *E* and *E′*. The point *D*, at which these rays of light seem to meet, is the apparent position of the bottom of the vessel. Due to the phenomenon of refraction, therefore, a vessel filled with water appears shallower than it really is.

On account of the refraction of the earth's atmosphere, the stars, with the exception of those directly overhead, appear displaced from their true positions. The light from a star is refracted as it enters the atmosphere, and this refraction increases as the ray penetrates the denser layers of the air which lie nearer the surface of the earth. The eye which receives this ray refers the star not to its true position *S*, Fig. 443, but to the apparent position *S′*, in the direction of the ray of light which enters the eye.

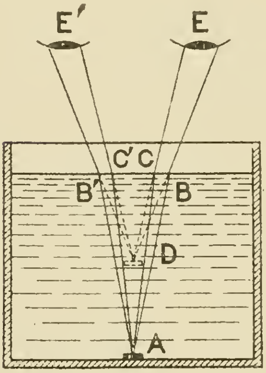

FIG. 442. The refraction of light causes water to appear shallower than it really is.

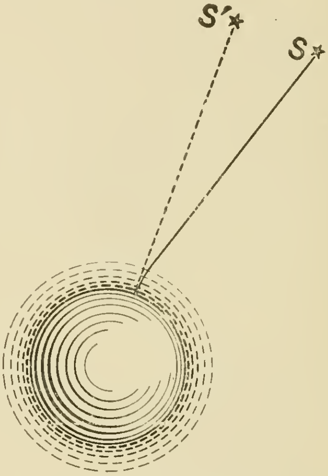

FIG. 443. The displacement of a star due to the refraction of light in the atmosphere.

When the setting sun has nearly touched the horizon, it is in reality entirely below the horizon, and has its apparent position still above this line, due to the refraction of the air. The oblate, or flattened, appearance which the sun presents at this time is due to the fact that the light from the lower edge passes through a denser layer of the earth's atmosphere than that from the upper edge, and therefore undergoes a greater refraction and a greater amount of deviation.

621. Refraction in a Medium Bounded by Plane Parallel Surfaces. When a ray of light passes through a medium having plane parallel sides, such as a piece of thick plate glass, the ray does not suffer angular deviation, but is displaced parallel to its original direction. Upon entering the refracting medium, Fig. 444, the light is bent towards the normal and proceeds in this direction until it encounters the second surface, when the angle of refraction at the first surface becomes the angle of incidence at the second, and the angle of refraction at the second surface is equal to the angle of incidence at the first. After passing obliquely through the glass, the ray therefore proceeds in a direction parallel to its original direction, but is slightly displaced laterally. The fact that a ray of light suffers no angular displacement in passing through a refracting medium bounded by plane parallel surfaces is of importance in a later demonstration.

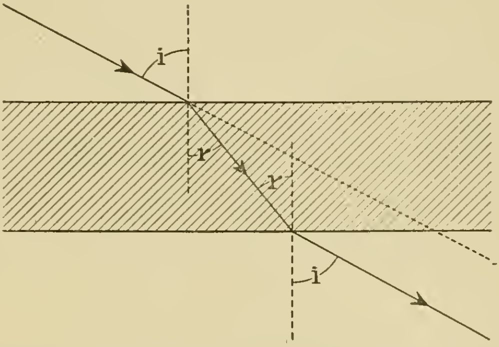

FIG. 444. Refraction of light in passing through a denser medium bounded by plane parallel surfaces.

622. Total Reflection—The Critical Angle. Let *mn*, Fig. 445, represent the boundary between a rare medium above and a denser one beneath, such as air and water, and let a ray of light pass from the air into the water at all possible angles of incidence. In each case the ray is so refracted that

$$\frac{\sin i}{\sin r} = n,$$

where n is the index of refraction from air to water. The ray NI, passing perpendicular to the surface, undergoes no deviation, but is propagated in the direction IN' in the denser medium. The ray AI is refracted in the direction IA', BI is refracted in the direction IB', CI is bent in the direction IC' and DI, which makes an angle of incidence of almost 90°, is refracted in the direction ID'.

Conversely, if light is sent from the denser into the rarer medium, that is, in this case from the water into the air, the incident ray $N'I$, perpendicular to the surface, passes out into the air undeviated in the direction IN. So, also, if light is sent from the water in the direction $A'I$, it is refracted and emerges in the direction IA. In the same way, if $B'I$ is the incident ray, IB is the refracted ray; if $C'I$ is the incident ray, IC is the refracted ray, and if light is sent from the denser into the rarer medium in the direction $D'I$, so that $D'I$ is an incident ray, the light just passes out into the air in the direction ID. It is evident that the incident ray $D'I$ in the denser medium is the last ray which can undergo refraction, for the refracted ray ID makes an angle of refraction of 90° in the rarer medium. The angle of incidence $N'ID'$ in the denser medium is therefore the greatest angle of incidence which permits the refracted ray to emerge in the rarer medium, and this angle of incidence is known as *the critical angle. When light passes from a denser to a rarer medium, the critical angle is that angle of incidence in the denser medium for which the angle of refraction in the rarer medium is 90°.*

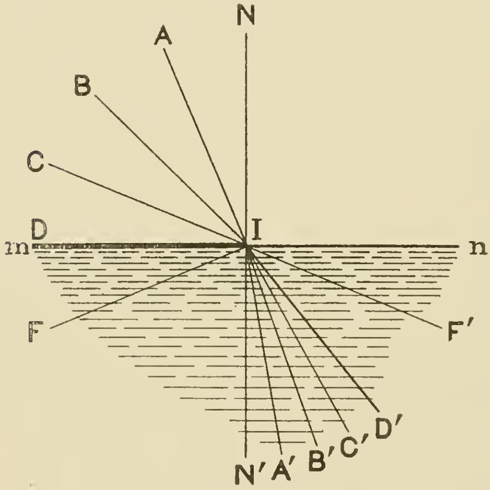

FIG. 445. The critical angle and the total reflection of light.

If $F'I$ represents a ray of light incident at an angle greater than the critical angle, this ray cannot emerge from the denser medium, and therefore undergoes total reflection at the point I, passing back, still in the denser medium, in the direction IF. *When light passes from a denser to a rarer medium, total reflec-*

tion occurs when the angle of incidence is greater than the critical angle.

We have seen that when light passes from a denser into a rarer medium,

$$\frac{\sin i}{\sin r} = \frac{1}{n},$$

where n is the index of refraction from the rarer to the denser medium. In this expression, if i is the critical angle, then r is 90°, and

$$\sin r = 1.$$

Accordingly, for the critical angle

$$\sin i = \frac{1}{n}.$$

That is, the sine of the critical angle is the reciprocal of the index of refraction; or, the critical angle may be defined as that angle of incidence in the denser medium, the sine of which is the reciprocal of the index of refraction.

623. Applications of Total Reflection. A piece of appara- known as the inverting prism is extensively used, especially with the projecting lantern, and depends in its action upon the phenomenon of the total reflection of light. This device is a right angled prism of glass, and is so placed before the lantern that rays of light must pass through it on their way to the screen. Let A and B, Fig. 446, represent two such rays. These rays enter the prism and are refracted, passing down to the points a and b respectively on the longest face of the prism. But the angle of the prism is such that the angle of incidence at this face is greater

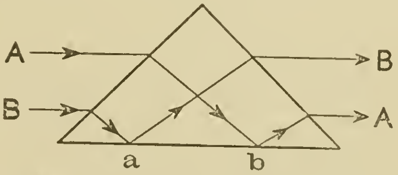

FIG. 446. A total reflecting or inverting prism.

than the critical angle, so that total reflection occurs; and accordingly, the rays after being again refracted, pass out parallel to their original directions. It will be observed, however, that the ray which was formerly above is now below, and *vice versa.* Since a lantern projects an object inverted, the total reflecting prism placed in the path of the ray inverts this already inverted image, and therefore causes it to appear erect upon the screen.

The experiment of the "illuminated vein" most beautifully illustrates the phenomenon of total reflection. A tall tank, Fig. 447, provided with a window on one side and an opening, closed by a cork, directly opposite, is filled with water. Sunlight or light from the electric lamp is concentrated by a lens before the window in the tank, so as to converge directly upon the opening on the other side. When the cork is removed, the water flows out in a parabolic path, and the light enters the stream, passing initially in the direction in which the water is moving. The light, however, strikes the sides of the jet very obliquely, and everywhere at an angle much greater than the critical angle. Accordingly, the light cannot escape, since it is totally reflected at every point of its path, and therefore remains in the stream to illuminate it throughout the entire course. The perfect parabolic path of the jet, illuminated from within by the light which everywhere undergoes total reflection, makes this experiment one of singular interest and beauty.

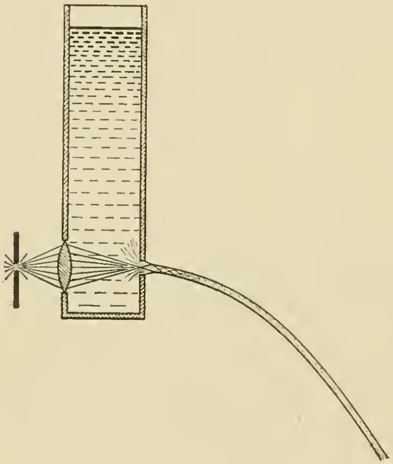

FIG. 447. The experiment of the "illuminated vein."

624. Refraction by a Prism. A prism in optics is, in general, any transparent body the sides of which are formed by three planes which meet in three parallel lines. The angle in which two of these planes meet is called *the refracting edge* of the prism, and the cross-section of the prism at right angles to the refracting edge is known as *the principal plane, or the principal section,* of the prism.

A ray of light which passes through a prism is refracted at both surfaces and undergoes *an angular deviation,* due to the fact that these surfaces are not parallel. The amount of this angular deviation, as well as the direction of the emerging ray, may best be understood by a diagram. Let the triangle, Fig. 448, represent the principal section of a prism made of some medium, like glass, which is denser than air, and let A be the refracting angle. A ray of light which is incident at I from a source S is refracted towards the normal $N I$, and passes through the prism in the direction $I I'$. At I' the ray again undergoes refraction, this time passing from the denser to the rarer medium, and hence is bent away from the normal, $N' I'$, passing out into the air in the direction $I' R$. The angle D, which is the angle between the original direction of the incident ray and the final direction of the refracted ray, is the angular deviation which has been produced, and is known as the *angle of deviation.*

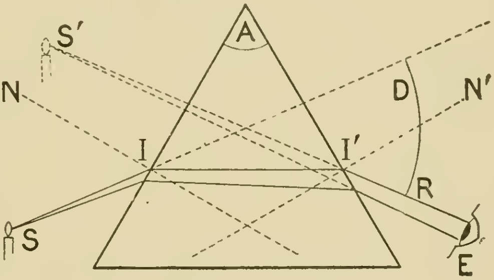

Fig. 448. Refraction of light by a prism.

If the prism is made of a material which is optically *denser* than the medium in which it is placed, *a ray of light in passing through it will be refracted or bent towards the thicker portion of the prism.* On the other hand, an eye which receives this refracted ray will perceive the source of light displaced towards the thinner portion, that is, towards the refracting edge. Indeed, if two rays from the source are followed, as these are refracted by the prism, they will be found to diverge after refraction, and hence they appear to meet at the point S', which is the virtual image of the source S. Thus, the image produced by a prism is virtual, is on the same side of the prism as the source of the light, and is always seen displaced towards the refracting edge.

625. The Position of Minimum Deviation. If the angle of incidence upon the first face of the prism is changed, the angle of refraction will also change. Consequently, the angle of incidence at the second face will then change, and the angle of refraction at this face will likewise change, so that the total angle of deviation, D, produced by the prism is seen to depend upon the angle of incidence at the first face. The manner in which the angle of deviation depends upon the first angle of incidence is, however, most interesting. If this angle of incidence is at first zero, and is then gradually and continuously increased, the angle of deviation, D, will be found to become smaller. Indeed, this angle will continue to grow smaller until a particular angle of incidence is attained, after which, for a further increase in the angle of incidence, the angle of deviation will not only not become smaller, but will actually grow larger. For a particular angle of incidence, therefore, and for this alone, the angle of deviation D attains a minimum value, and the prism is then said to be in its position of *minimum deviation.*

One further matter of interest and importance is connected with the position of minimum deviation, namely, the fact that in this position the ray of light then passes symmetrically through the prism. That is, if D, Fig. 449, represents the angle of minimum deviation of the ray $S I$ from the source S, *the ray* $I I'$, in passing through the prism *makes equal angles with each face, and the angle of incidence,* i, *at the first face is equal to the angle of refraction,* r', *at the second.* Accordingly, when the prism is in the the position of minimum deviation, it follows that either an increase or a decrease in the angle of incidence will produce an increase in the angle of deviation, D, and also that the ray of light then passes symmetrically through the prism, the angles made by the ray with one face being equal to the corresponding angles made by the ray with the other face.

Whenever a prism is used as a means of refracting light, it is always to be placed in its position of minimum deviation. This may best be accomplished by maintaining the ray in its original direction, and varying the angle of incidence by turning the prism. It is therefore usual to receive the ray either by the eye or upon a screen, and then with the hand to rotate the prism about its axis until the angle of deviation is observed to increase, whichever way the prism is turned. The prism is then in its position of minimum deviation, and the ray of light passes symmetrically through it.

626. Determination of the Index of Refraction by a Prism. Because of the ease and accuracy of determining the angle of deviation produced in a ray of light by a prism, this body is the means usually employed in measuring the index of refraction of the medium of which the prism is made. This may be accomplished in the following manner.

Let *SI*, Fig. 449, be a ray of light incident upon the prism in the plane of the principal section, *SIN* the angle of incidence *i*, and *MII′* the angle of refraction *r*. The ray of light *II′* within the prism makes the angle of incidence *II′M* or *i′*, upon the second face of the prism, and emerges in the direction *I′R*, making the angle of refraction *RI′N′*, or *r′*, at the second surface. If the directions of the incident and the refracted rays are produced until they meet at *O*, the angle *D*, which is the angle included between these lines, is the total angle of deviation which has been produced by the prism. Furthermore, the angle *d* is the angle of deviation produced by the refraction at the first surface, and *d′* is the angle of deviation produced by the refraction at the second surface, so that

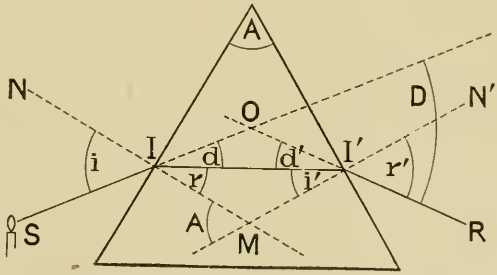

FIG. 449. In the position of minimum deviation, a ray of light passes symmetrically through the prism.

$$D=d+d'.$$

But

$$d=i-r.$$

and

$$d'=r'-i',$$

therefore

$$D = i + r' - (r + i').$$

But

$$r + i' = A,$$

(the exterior angle of a triangle is equal to the sum of the two opposite interior angles), where A is equal in magnitude to the angle of the prism (their sides being perpendicular).

If now the prism is in the position of minimum deviation, that is, if the ray of light passes symmetrically through it,

$$r' = i,$$

and

$$i' = r.$$

Accordingly,

$$A = 2r,$$

and

$$r = \frac{A}{2}.$$

Therefore, in the position of minimum deviation,

$$D = 2i - A,$$

from which

$$i = \frac{D + A}{2}.$$

But we know (p. 777) that

$$\frac{\sin i}{\sin r} = n,$$

where n is the index of refraction. When we substitute in this equation the above values of i and r, thereby applying this expression to the refraction of light produced by a prism, we obtain

$$\frac{\sin i}{\sin r} = n = \frac{\sin \frac{D + A}{2}}{\sin \frac{A}{2}}.$$

Since the angle of the prism A and the angle of deviation D may be determined with great accuracy, the prism is the best means of measuring the index of refraction, and this is the method ordinarily employed.

In a similar way, indices of refraction of liquids may be determined by filling a hollow prism with the liquid to be investigated. The plane parallel glass surfaces of the hollow prism produce no angular deviation of the ray, hence the deviation obtained is that due solely to the prismatic volume of the liquid contained.

627. Refraction at a Spherical Surface. Let AB, Fig. 450, be a portion of a spherical surface separating the denser medium to the right from the rarer medium to the left, and let C be the center of curvature of this surface. Let L be a source of light upon the principal axis LA, and let LB be a ray of light incident at B. This ray of light will be refracted towards the normal in the denser medium, and the refracted ray, if produced backwards, will cut the axis in some point such as l, so that l is the virtual focus conjugate to L. Accordingly, the angle LBC is the angle of incidence i, and the angle lBC is the angle of refraction r.

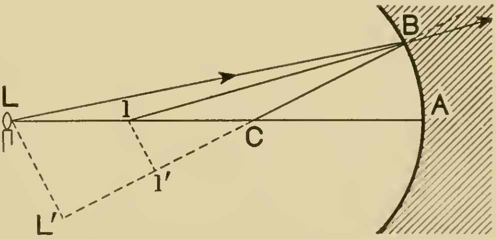

FIG. 450. Refraction at a spherical surface.

If n is the index of refraction of the denser medium, we know that

$$n = \frac{\sin i}{\sin r} = \frac{\sin L B C}{\sin l B C}.$$

In order to determine the values of these sines, we must drop the perpendiculars LL' and ll' from the points L and l respectively upon the line CB produced. Then

$$\sin L B C = \frac{L L'}{L B},$$

and

$$\sin l B C = \frac{l l'}{l B}.$$

The above expression for the index of refraction then becomes

$$n = \frac{\frac{L L'}{L B}}{\frac{l l'}{l B}},$$

or

$$n = \frac{L L'}{l l'} \cdot \frac{l B}{L B}.$$

But from similar triangles,

$$\frac{L L'}{l l'} = \frac{L C}{l C},$$

and

$$\frac{l B}{L B} = \frac{l A}{L A} \text{ (approximately).}$$

Therefore,

$$n = \frac{L C}{l C} \cdot \frac{l A}{L A}.$$

Letting

$$LA=p,$$

$$lA=p',$$

$$CA=r,$$

this equation becomes

$$n=\frac{p-r}{p'-r}\cdot\frac{p'}{p}.$$

Clearing of fractions,

$$npp'-npr=pp'-p'r,$$

or

$$pp'\ (n-1)=npr-p'r.$$

Dividing by $pp'r$,

$$\frac{n}{p'}-\frac{1}{p}=\frac{n-1}{r}\ \cdots\cdots\cdots\ \text{(A)},$$

which is the fundamental equation for the refraction of light at the spherical surface of a denser medium.

In this connection it is interesting to notice that if

$$n=-1,$$

that is, if the light is reflected at the spherical surface back again into the medium at the left, equation (A) then becomes

$$\frac{1}{p}+\frac{1}{p'}=\frac{2}{r},$$

which is the fundamental expression for reflection from a concave surface (p. 761). This, however, is not further considered, since we are dealing with refraction in the second medium.

628. Lenses. If the denser medium is terminated by a second surface, either plane or spherical, it constitutes what is known as a lens. A lens, therefore, is a portion of a transparent refracting substance bounded by two spherical surfaces, or by two surfaces, one spherical and the other plane. The ray of light therefore undergoes refraction at this second surface also, and so re-enters the original medium, having passed through the lens. In this case, however, the index of refraction for this second surface is $\frac{1}{n}$ since the light passes from the denser to the rarer medium. Moreover, the light is incident upon this second surface as if it came from the virtual focus l, Fig. 451, and hence, if the lens is so thin that we may neglect its thickness, p' becomes the distance of the source of light from the second surface. Furthermore, we will let q be the distance from this second surface at which the refracted ray crosses the axis, that is, q will be the distance of the focus from the second surface, and r' will represent the radius of curvature of the second surface. Accordingly the equation for the second surface which corresponds to equation (A) for the first becomes

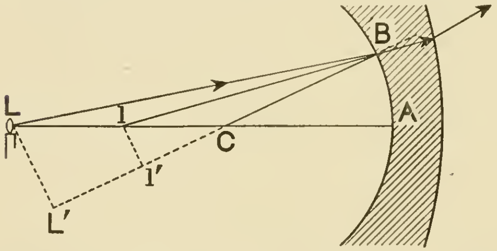

FIG. 451. Refraction at the two surfaces of a lens.

$$\frac{\frac{1}{n}}{q} - \frac{1}{p'} = \frac{\frac{1}{n} - 1}{r'},$$

from which

$$\frac{1}{n\,q} - \frac{1}{p'} = \frac{1 - n}{n\,r'},$$

or

$$\frac{1}{q} - \frac{n}{p'} = \frac{1-n}{r'}, \quad \text{. . . . .} \quad (B)$$

which is the fundamental equation for refraction at the second spherical surface.

Since the light passes entirely through the lens, undergoing refraction at both surfaces, the effect of these two refractions may be obtained by adding equations (A) and (B). Accordingly,

$$\frac{1}{q} - \frac{1}{p} = (n-1)\left(\frac{1}{r} - \frac{1}{r'}\right).$$

This may be called the fundamental equation of the lens.

Since p is the distance of the source of light from the lens, and q is the distance of the focus from the lens, p and q are at conjugate foci. Moreover, since the thickness of the lens has been neglected, this expression is true only for very thin lenses, that is, when the thickness of the lens is vanishingly small as compared with the distances p and q.

When the source of light is at infinity, the point where the rays meet after refraction is known as the principal focus. The distance of the principal focus from the lens is called the principal focal length of the lens, and this value of q is denoted by f. Therefore, when

$$p = \infty,$$

the equation for the lens becomes

$$\frac{1}{q} = (n-1)\left(\frac{1}{r} - \frac{1}{r'}\right).$$

or

$$\frac{1}{f} = (n-1)\left(\frac{1}{r} - \frac{1}{r'}\right).$$

The convention which has here been adopted as to signs is to regard all distances on the side of the lens towards the source of light as positive, and all distances from the lens in the opposite direction as negative. Since n is larger than unity, $(n-1)$ is always a positive quantity, and therefore f, which is the distance of the principal focus from the lens, must depend for its sign upon the expression

$$\frac{1}{r} - \frac{1}{r'},$$

Here it must be remembered that r and r' are the radii of curvature of the first and the second surfaces respectively, and that these distances may be positive or negative according to the shape of the lens. It will be seen, Fig. 452, that six forms of lenses are possible, depending upon the values and the signs which r and r' may have. The first three are thicker in the center than at the edges, and together form lenses of one class, while the second three are thinner at the center than at the edges, and these form lenses of another class.

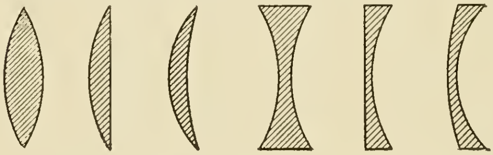

FIG. 452. The six possible forms of lenses.

If we draw radii of these surfaces, and test the sign of the expression

$$\frac{1}{r} - \frac{1}{r'},$$

we shall then be able to determine the sign of f for each lens. Making this application, for the sake of brevity, to the first lens only, Fig. 453, we see that r, the radius of the first surface is, negative, and therefore $\frac{1}{r}$ is also negative. r', the radius of the second surface, is positive, and $\frac{1}{r'}$ is also positive. $\frac{1}{r'}$, however, has the minus sign before it, and therefore the expression

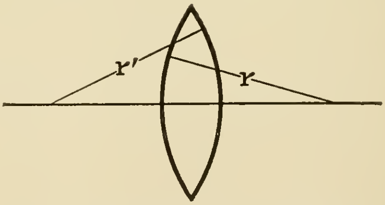

FIG. 453. Determination of the positive or negative value of the expression.

$$\frac{1}{r} - \frac{1}{r'},$$

for the first lens is negative, and f is negative. Applying this test to each of the six lenses in turn, we find that f is negative for the first three, and positive for the second three. The negative focal length in the case of the first three lenses indicates that parallel rays of light after passing through the lens really come to a focus on the farther side, and these three lenses are therefore known as converging lenses. The positive focal length in the case of the second three lenses indicates that parallel rays of light, after passing through the lens, continue to diverge, for the principal focal distance then falls on the side of the lens towards the source of light, and this focus is therefore virtual. On this account, the second three lenses are called diverging lenses, and always give virtual foci.

Moreover, q is always negative for real foci, since the rays of light, if they meet after passing through the lens, must do so on the side of the lens away from the source. A focus, therefore, is real when it is negative.

Inasmuch as

$$\frac{1}{f} = (n - 1)\ (\frac{1}{r} - \frac{1}{r'}),$$

we see that the second member of this equation is the same as the second member of the fundamental equation of the lens (p. 793). Equating the first members, therefore, we obtain

$$\frac{1}{q} - \frac{1}{p} = \frac{1}{f}.$$

Here f and q, for converging lenses, are essentially negative, and their positive values indicate that these distances are to be taken on the side of the lens away from the source of light. For diverging lenses, f and q are inherently positive, and are always found on the positive side of the lens.

629. Conjugate Foci in Converging Lenses. The equation last derived is extremely convenient in practice as showing the relation between conjugate foci and the focal length of the lens. The relative movement of the conjugate foci may be beautifully shown by allowing p to occupy different position before the lens.

1. If

$$p = \infty,$$

this equation becomes

$$\frac{1}{q} = \frac{1}{f},$$

or

$$q = f.$$

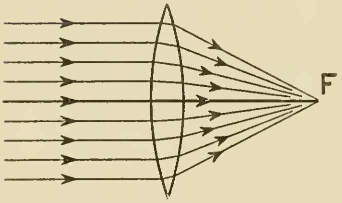

FIG. 454. When the source of light is at infinity, the rays meet at the principal focus.

Therefore, when the source of light is at infinity, that is, when the rays of light come parallel to the principal axis, Fig. 454, the light is brought to a focus at the principal focus.

2. If

$$p = -2f,$$

that is, if the source of light is at a distance from the lens equal to twice the focal length and on the positive side ($-2f$ is positive, since f itself is negative), then

$$\frac{1}{2f} + \frac{1}{q} = \frac{1}{f},$$

from which

$$\frac{1}{q} = \frac{1}{2f},$$

and

$$q=2f.$$

Therefore, if the source of light is at a distance from the lens equal to twice its focal length, Fig. 455, the focus conjugate with this source is likewise at a distance from the lens equal to twice its focal length. The source and its focus are, then, four times the focal length of the lens apart, which is the shortest distance that can exist between a source and its real focus.

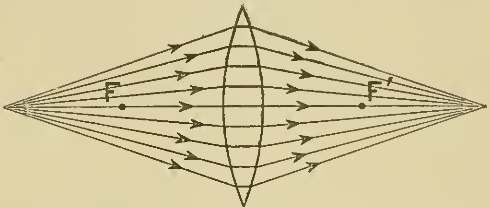

FIG. 455. When the distance of the source from the lens is twice the focal length, the distance of the focus from the lens is twice the focal length.

3. If,

$$p=-f,$$

that is, if the source of light is at the principal focus, the equation for the location of the conjugate foci becomes

$$\frac{1}{f}+\frac{1}{q}=\frac{1}{f},$$

from which

$$\frac{1}{q}=0,$$

and

$$q=\infty.$$

Therefore, when the source of light is at the principal focus, Fig. 456, the rays of light, after passing through the lens, leave it parallel to the principal axis, that is, these rays meet at infinity.

4. When the source of light is between the principal focus and the lens, that is, when

$$p<-f,$$

then

$$\frac{1}{q} > -\frac{1}{f},$$

or

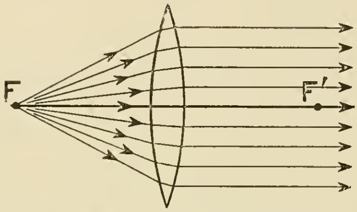

FIG. 456. When the source of light is at the principal focus, the rays leave the lens parallel to the principal axis.

$$-\frac{1}{p} > \frac{1}{f},$$

and the above equation becomes

$$\frac{1}{q} + \frac{1}{p} = \frac{1}{f},$$

from which

$$\frac{1}{q} = \frac{1}{f} - \frac{1}{p},$$

where

$$-\frac{1}{p} > \frac{1}{f}.$$

That is, $\frac{1}{q}$ equals $\frac{1}{f}$ minus a quantity which is greater than $\frac{1}{f}$, and the right side of this equation is negative. But, since q itself is essentially negative, we have here the negative value of a negative quantity, or, in other words, a positive quantity. Accordingly, when the source of light is between the principal focus and the lens, Fig. 457, the rays do not meet on the farther side of the lens, but continue to diverge after refraction, and these rays produced backward seem to meet at a point on the same side of the lens as the source, and therefore at a virtual focus. Accordingly, light is brought to a real focus only when the source of light is at a distance from the lens greater than the principal focal length.

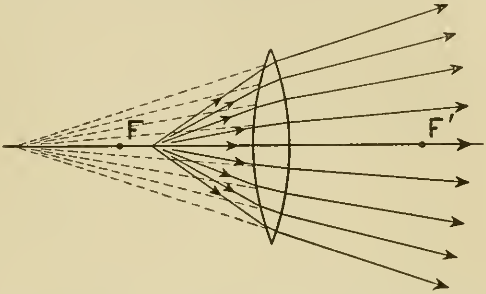

FIG. 457. When the source is nearer the lens than the principal focus, the rays continue to diverge, and the focus is virtual.

630. The Optical Center of a Lens. The optical center of a lens is defined as the point in a lens or near it such that all rays of light passing through this point suffer no angular deviation by the lens. That this point is fixed with respect to the lens is shown as follows:

Let COC' Fig. 458, be the principal axis of the lens, C and C' being the centers of curvature of the spherical surfaces. Draw two parallel radii such as CI and $C'I'$, and let the line II' represent the direction of the ray of light within the lens. Since this ray of light makes equal angles with the normals CI and $C'I'$ within the lens, that is, since the angle of refraction CII' at the first surface is equal to the angle of incidence $C'I'I$ upon the second surface (alternate interior angles), the angles of incidence and emergence of the ray must be equal. Therefore, the ray of light SI which passes in the direction II' through the lens emerges in the direction $I'R$, parallel to its original direction SI. If tangent planes are erected at I and I' these planes are parallel, since they are perpendicular respectively to parallel radii. Accordingly, the ray $SII'R$ passes through the lens as if

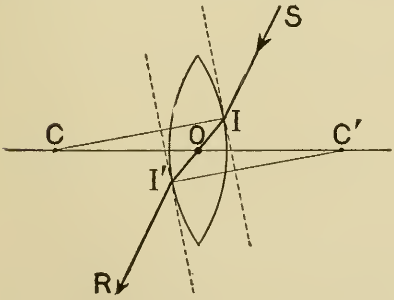

FIG. 458. Diagram showing that the optical center O is a fixed point.

it were passing through a medium bounded by plane parallel surfaces, and therefore suffers no angular deviation. It is true that there is a lateral displacement, but if the lens is thin, this may be neglected. Let O be the point where the ray II' intersects the principal axis CC'. We are to show that every ray passing through O suffers no angular deviation, or, in other words, that the point O is constant.

In the similar triangles COI and $C'OI'$,

$$\frac{C\,O}{C\,I} = \frac{C'\,O}{C'\,I'},$$

or

$$\frac{C\,O}{C'\,O} = \frac{C\,I}{C'\,I'};$$

but

$$\frac{C\,I}{C'\,I'} = \frac{r'}{r} = \text{constant},$$

therefore

$$\frac{C\,O}{C'\,O} = \text{constant},$$

from which it follows that O is a constant point, since C and C' are fixed.

Accordingly, every ray which passes through the point O suffers no angular deviation in passing through the lens. Moreover, since we are dealing only with thin lenses, we may say that any ray passing through the point O passes straight through the lens without deviation. This point O is known as the *optical center* of the lens. For a double convex or a double concave lens, the optical center is within the lens; for a plano-convex or a plano-concave lens it is at the spherical surface; and for a lens of either of the other two forms, it is outside the lens, near the surface of the shorter radius of curvature.

631. Axes. *An axis of a lens is any straight line passing through the optical center. The particular axis passing through the centers of curvature is called the principal axis, while all other axes are known as secondary axes.* Since a lens is a solid of revolution, the principal axis is also the axis of revolution. The expressions which apply to conjugate foci on the principal axis apply also to conjugate foci on secondary axes, provided the latter make only a small angle with the principal axis.

632. The Formation of Images by Lenses. Rays of light which depart from a luminous object placed before a lens pass through the lens, and produce an image of the object at the point where these rays meet after refraction. Although the number of rays thus passing through a lens is very great, two especially may be followed without difficult computation. Of the infinite number of rays which depart from any point of an object, that particular one which passes parallel to the principal axis will be so refracted as to pass through the principal focus. Accordingly, the image of the point will be found somewhere upon this latter ray. Also that particular ray which passes through the optical center of the lens continues in its original direction without undergoing angular deviation, and the image of the point will be found likewise somewhere upon this ray. Since the image of the given point is thus found upon both of these rays, it must be found at their point of intersection. In like manner the images of other points of the object may be located; and, since the image of an object is the assemblage of the images of its points, this image may be easily constructed.

Inasmuch as lenses may be divided into two general classes, converging and diverging, the double convex and the double concave lenses will be taken as types of these classes respectively.

633. Images in Converging Lenses. In the formation of images in converging lenses, two cases are to be distinguished, depending upon the position of the object with relation to the principal focus of the lens.

1. The object is farther from the lens than the principal focal distance. Let *XY*, Fig. 459, be an object before a double convex

lens at a distance from the lens greater than the principal focal length. From the point *X* one ray, and only one ray, *XA*, will pass parallel to the principal axis, and this will be refracted through the principal focus *F′*. Therefore, somewhere on the line *AF′* will be found the image of the point *X*. Also from *X* one ray, and only one ray, *XO*, will fall upon the lens so as to pass through its optical center *O*, and this ray will pass through the lens undeviated. Therefore, somewhere on the line *XO* will be found the image of the point *X*. Since the image of this point is found on the line *AF′* and also on the line *XO*, it must be found at the point of intersection *x*. *x*, therefore, is the image of the point *X*. In the same way, the ray *YB*, falling parallel to the principal axis, is refracted through the principal focus *F′*, and therefore the image of *Y* is found somewhere on the line *BF′*. So, also the ray *YO*, passing through the optical center *O*, will continue undeviated, and the image of *Y* will be found somewhere on the line *YO*. Since this image is found on the line *BF′*, and also on the line *YO*, it must be located at their point of intersection *y*. *xy*, therefore, is the image of the object *XY*, and is real and inverted.

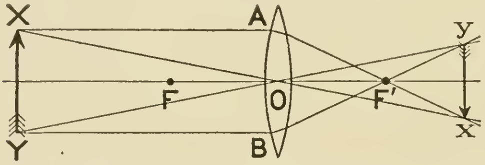

FIG. 459. The formation of an image in a converging lens when the object is beyond the principal focus.

The relative size of object and image may be readily determined. Since the triangles *XOY* and *xOy* are similar,

$$\frac{XY}{xy} = \frac{p}{q}$$

where *p* and *q* are the distances of the object and image from the lens respectively. Therefore, *the relative sizes of an object and its image are directly as their respective distances from the lens.* When

$$p = -2f,$$

we have seen (p. 796) that likewise

$$q=2f,$$

that is, when the object is at the distance from the lens equal to twice its focal length, the image is also at a distance from the lens equal to twice its focal length, and the object and image are then of the same size. A convenient method, therefore, of measuring the focal length of a lens is to adjust the object before the lens until the object and image are of the same size. The object and its image are then four times the focal length of the lens apart.

2. The object is between the principal focus and the lens. In this case, precisely the same construction is to be followed as before. One ray from the point X of the object, Fig. 460, will fall parallel to the principal axis, and this will be refracted so as to pass through the principal focus F'. Therefore, somewhere on the line AF', or on this line produced, will be found the image of the point X. Moreover, one ray XO will fall upon the lens so as to pass through the optical center O, and this ray will be undeviated by the lens, hence somewhere on the line XO, or on this line produced, will be found the image of the point X. Since the image of this point is found both on the line AF' and also on the line XO, it must lie at their point of intersection. Inasmuch as these rays continue to diverge after passing through the lens, they must be produced backwards, and the point x where they apparently meet is the image of the point X. In a similar way, the point y is located as the image of the point Y. xy, therefore, is the image of the object XY, and is erect, virtual, and magnified.

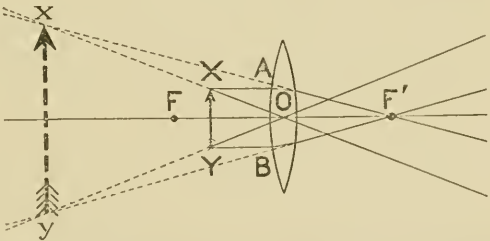

Fig. 460. The formation of an image in a converging lens when the object is between the principal focus and the lens.

Here again the relative sizes of object and image are as the respective distances of the object and image from the lens. This magnified, erect, and virtual image of the object is the image produced by the ordinary magnifying glass.

634. **The Image in a Diverging Lens.** In a diverging lens, the image is formed in every way analogous to its construction in a converging lens. F and F' are the virtual principal foci of a double concave lens, Fig. 461, and O is its optical center. From the point X of the object, the ray XA falls parallel to the principal axis, and is refracted by the lens as if it came from the principal focus F. Therefore, somewhere on the line FA will be found the image of the point X. Also from the point X, the ray XO falls upon the lens so as to pass through the optical center, and continues through the lens undeviated. The image of the point X must lie, therefore, somewhere on the line XO. Since the image of this point occurs on both the lines FA and XO, it must be found at their point of intersection x. In the same way, y is located as the image of the point Y, and these two points define the position of the image xy.

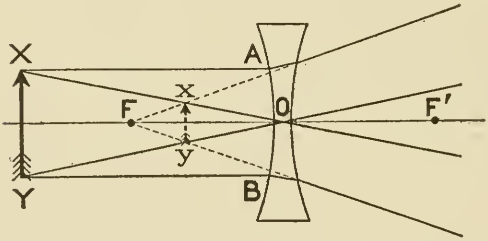

Fig. 461. The formation of an image in a diverging lens. In this case the image is always virtual.

From the similar triangles XOY and xOy, it will be seen that the relative sizes of the object and image are as their respective distances from the lens. For all positions of the object, the image in a concave lens is erect, virtual, and smaller than the object.

635. **Spherical Aberration of Lenses.** When light falls upon a large lens or one of strong curvature, the rays which are incident near the edges strike the surface under a greater angle of incidence than do those which fall upon the central portions. Accordingly, rays passing through the edges of a lens, Fig. 462, undergo a greater refraction, and therefore come to a focus nearer the lens, than rays passing through the central portions. In a lens with perfectly spherical surfaces, therefore, the focus is not a sharp point, but is an elongated cusped curve, Fig. 463, called the *caustic*, formed by the refracted rays crossing each other two and two.

This failure of a lens to bring rays of light exactly to one focus is called *spherical aberration,* and produces a blurring, or indistinctness, of the resulting image. Spherical aberration may be corrected by changing the forms of the refracting surfaces; and while this is extremely difficult to do with accuracy, it is, nevertheless, the method employed in making object glasses for large telescopes. The forms of the surfaces of these lenses are accordingly changed, usually by hand, until the resulting image is perfectly sharp. For less accurate purposes, however, it is customary to interpose a diaphragm or screen which cuts off the rays which pass through the edges of the lens, so that those which come through the center may be brought more nearly to a sharp focus. This method has the disadvantage of diminishing the amount of light passing through the lens, and therefore of decreasing the brilliancy of the image produced.

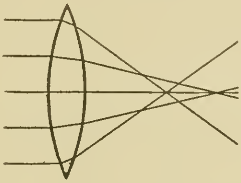
FIG. 462. The cause of spherical aberration in a lens.

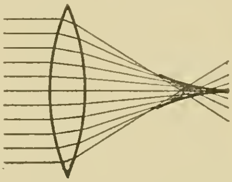
FIG. 463. The caustic by refraction.

636. The Eye. The eye, considered as an optical instrument, consists of several parts, one of which is a strongly convex portion called the cornea, which is filled with the clear, transparent, aqueous humor *A*, Fig. 464, and behind this is the dense, highly-refracting crystalline lens *L*. Before this lens is a circular screen, called the iris, which opens and closes involuntarily, thus admitting more or less light into the eye. The space behind the crystalline lens is filled with a dense liquid, *V*, called the vitreous humor. At the rear of the eye is an extremely delicate and sensitive membrane, known as the retina, which is, indeed, only an expansion of the optic nerve, and which serves as the screen upon which the real, inverted image of the object is thrown by the crystalline lens.

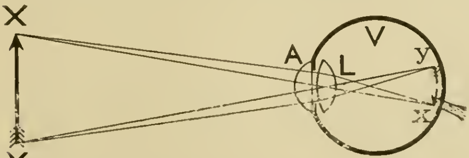

FIG. 464. The media of the eye, and the formation of an image upon the retina.

Since an object and its image are always at conjugate foci, it is seen that if the crystalline lens were rigid, the image would fall exactly upon the retina only when objects were at one particular distance away, and therefore only objects at this distance would appear sharply defined. It is a remarkable fact, however, that the eye has the power of accommodating itself to different distances by changing the convexity of the crystalline lens. As the object moves with reference to the eye, the crystalline lens changes its curvature, and always in such a way as to throw a sharp image upon the retina. The convexity of the lens increases as the object approaches the eye, until the distance between the two is less than the distance of distinct vision, which in normal eyes is about ten inches. If the object is nearer the eye than this distance, the image is thrown too far back to be sharply defined, and the strain upon the eye which is then felt is the effort which is being made to change the curvature of the crystalline lens sufficiently to bring the image upon the retina.

637. Some Imperfections of the Eye and Their Correction. While in normal eyes the crystalline lens is of the proper curvature to focus the image of an object sharply upon the membrane of the retina, in near-sighted eyes this lens is too convex, Fig. 465, so that the image of the object falls between the lens and this membrane, thus producing an indistinctness of the image.

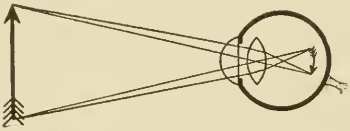

Fig. 465. In a nearsighted eye the image is formed in front of the retina.

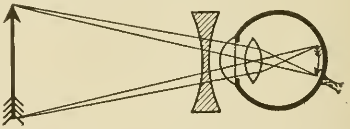

Fig. 466. The correction of a near-sighted eye by a concave spectacle lens.

From the movements of the conjugate foci, we know that as the object moves towards the lens, the image moves away from it. Therefore, in a near-sighted eye, the object must be held nearer the eye than the usual distance of distinct vision, in order that the image may be thrown sufficiently far back. A near-sighted eye can therefore be aided by a concave spectacle lens, which causes a divergence of the rays of light, Fig. 466, thus enabling the crystalline lens to throw the image far enough back to fall upon the retina. In far-sighted eyes the crystalline lens is too flat, Fig. 467, and therefore the image is thrown farther back than the rear of the eye. It is accordingly necessary, in a far-sighted eye, to employ a convex spectacle lens, which increases the convergence of the light, Fig. 468, and therefore enables the image of the object to be sharply defined upon the retina.

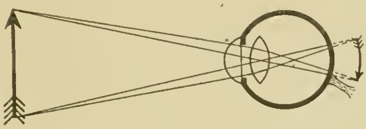

FIG. 467. In a far-sighted eye the image falls beyond the retina.

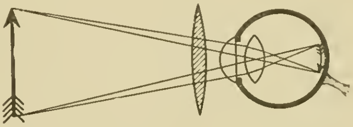

FIG. 468. The correction of a far-sighted eye by a convex spectacle lens.

In many eyes the curvature of the lens is greater in one plane than in a plane at right angles to this, so that light in the first plane is brought to a focus nearer the lens than that in the second. If the convexity in the vertical plane, for instance, is greater than that in the horizontal it follows that light from vertical objects will be brought to a focus nearer the lens than will light from horizontal objects. Accordingly, if horizontal lines are sharply defined, vertical lines will be indistinct, since they are brought to a focus before the retina is reached. This produces a result known as *astigmatism*. To correct for this defect, spectacles which are slightly cylindrical must be used, the strongest curved portion of the lens being placed opposite that plane of the eye where the curvature of the crystalline lens is weakest, and *vice versa*. The more strongly convex portion of the spectacle corrects the more slightly curved portion of the crystalline lens, and therefore an image sharply defined in all planes is thrown upon the retina.

638. After Images. When the image of an object falls upon the retina of the eye, the nerves of this sensitive membrane are stimulated, and transmit an impression of the object to the brain. That this process ultimately produces a fatigue of the retina may be perceived by gazing attentively for two or three minutes at a brightly illuminated diagram or picture possessing strong contrast, such as the heavy black lines upon the white page, shown in Fig. 469, and then closing the eyes or, better, turning them at once towards a white wall. In a few moments an effect, known as the *after image,* will be perceived, which is, indeed an image of the object, but with the white and black portions reversed, that is, the parts which were black upon a white field now appear light upon a dark field. Similarly, if a piece of brightly colored paper, lying upon a white surface, is observed with fixed attention, and the paper is then suddenly removed, an after image will appear for an instant in a color which is complimentary to that of the paper itself. Thus, if red paper is observed in this way, the after image will appear green; if yellow paper is observed, the after image will appear blue, and so on.

This shows that the retina has become fatigued by the constant perception of the object, and that the effort made by the reaction of the nerves is to restore this membrane to its normal condition.

639. Persistence of Vision. While the eye receives a sensation almost instantly, the impression so made lasts for an appreciable time, as may be perceived when a stick with a lighted coal at the end is whirled about the head at night. The impression made by the coal in every position persists until an entire revolution of the stick has been made, and the single coal has the appearance of being an illuminated ring.

The kinetoscope consists of a series of photographs of a moving object which are thrown in rapid succession upon a screen. The eye retains the impression made by one until the next appears, and the sensation produced is that of the moving object itself. This ability of the eye thus to retain for some time an impression made upon it is known as the *persistence of vision.*

640. Optical Illusions. Although the eye is remarkable for its range of vision and quickness of perception—even the form of a flash of lightning is distinctly observed—this organ is not always infallible in the impression of an object which it transmits to the brain. Thus, the rising full moon appears much larger near the horizon than higher in the sky, since in its lower position we are misled by our standards of experience, and unconsciously compare it with familiar objects at a distance which are known to be of large dimensions.

The series of parallel vertical lines crossed by the short diagonals, shown in Fig. 469, illustrates the false impression which the eye may give of an object which it correctly perceives. Careful observation will show that the short lines are straight and continuous across the vertical ones, although they seem to be displaced upon the two sides, the short lines on one side appearing to be prolongations of the spaces between these lines on the other. More immediately obvious, however, is the inclined appearance of the vertical lines, although, in reality they are strictly parallel. This deception is occasioned, in part at least, by the involuntary and inevitable wandering of the eye to parts adjacent to those which are directly observed, so that a bias is given to the eye by the inclined directions of the shorter lines. That this is so may be proved

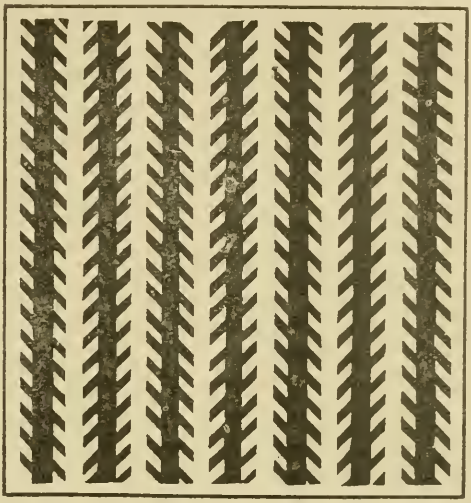
FIG. 469. The short diagonal lines seem displaced, and the vertical parallel lines appear inclined.

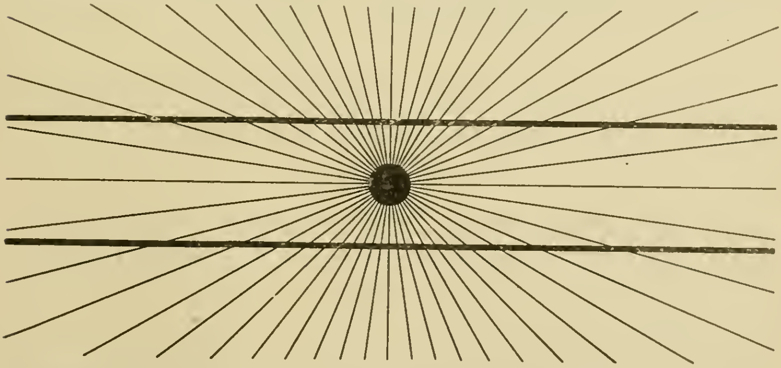
FIG. 470. By an optical illusion, the horizontal parallel lines are apparently distorted.

by obtaining the after image of this diagram in the manner already described, when the vertical lines will appear indeed vertical, without the slightest trace of inclination, thus showing that the eye had really received the correct impression, although the interpretation which it had placed upon this impression had been false. To a similar cause may be attributed the apparent distortion of the two horizontal parallel lines, shown in Fig. 470.

False impressions of objects observed, such as these and many others, form a group of phenomena known as *optical illusions*.

641. The Visual Angle. By the *visual angle* is meant the angle which an object subtends at the eye, and this varies inversely as the distance, Fig. 471. Accordingly, our estimate of the size of an object is based upon the visual angle which the object thus includes. It is evident that a large object will subtend a greater angle than a small one when both are at the same distance from the eye. It is also evident that an object of a given size when at one distance will subtend a certain visual angle, and that an object of twice this size at twice the distance will subtend only the same visual angle. A larger visual angle, and therefore greater accuracy in the perception of details, is obtained when the object is held near the eye, so long as the distance from the object to the eye is not less than the limit of distinct vision.

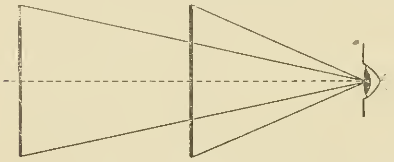

FIG. 471. The visual angle which an object subtends varies inversely as the distance of the object from the eye.

642. Optical Instruments—The Simple Microscope, or Magnifying Glass. The purpose of many optical instruments is to increase the visual angle by which the image of an object is seen. Thus, in the simple magnifying glass, Fig. 472, the glass is brought before the object XY, so that the latter is *between the principal focus and the lens,* after which it is further adjusted so that the virtual, erect, and magnified image xy of the object is formed at the distance of distinct vision, which is usually about ten or twelve inches from the eye. The magnifying power of the

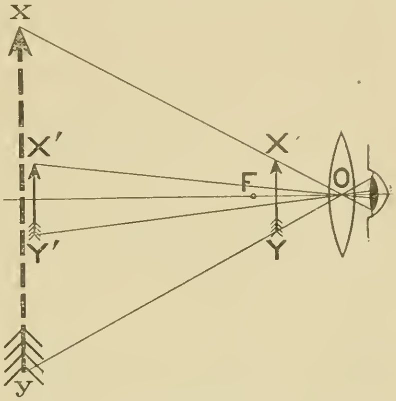

Fig. 472. Diagram of the magnifying power of the simple microscope.

simple magnifying glass, therefore, is the ratio of the angle which the image at the distance of distinct vision subtends at the eye to the angle which the object would subtend if it were placed at the distance of distinct vision. That is, the image subtends the angle xOy, while the object XY, if it were at the distance of distinct vision, and therefore in the position $X'\ Y'$, would subtend the angle $X'OY'$. Accordingly the ratio of the angle which the image subtends at the eye to that which the object would subtend at the distance of distinct vision is

$$\frac{x\,O\,y}{X'\,O\,Y'}$$

which is the magnifying power of the lens. Inasmuch as $X'Y'$ equals XY, the ratio of the above angles is the same as the ratio

$$\frac{O\,x}{O\,X},$$

and therefore very nearly equal to the ratio

$$\frac{O\,x}{O\,F},$$

since the object XY is brought between the lens and the principal focus, and very near to the latter point. The magnifying power of the simple microscope, or magnifying glass, therefore, is practically the ratio of the distance of distinct vision to the focal length of the lens.

643. The Compound Microscope. In the compound microscope, Fig. 473, the small, short-focus lens *AB*, called the objective, is brought near the object *XY*, so that the latter is *just outside the principal focus F*. Since the object is just beyond the principal focus and very near to it, a real, inverted, and magnified image of the object is formed at *xy* toward the upper part of the tube. A second lens *CD*, called the eye-piece, is now inserted above the image, and this is used as a simple magnifying glass with which to view the image *xy*. That is, the eye-piece *CD* is so adjusted *that the image xy is just within its principal focus F′*, and therefore a virtual, erect, and magnified image of this image is formed at *x′y′*. The angle which this image subtends at the eye is *x′Oy′*, while the angle which the object subtends at the eye is *XOY*. The ratio of these angles, that is,

$$\frac{x'Oy'}{XOY},$$

is the magnifying power of the compound miscroscope.

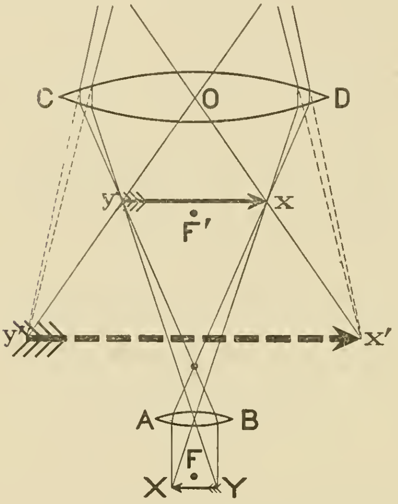

Fig. 473. Diagram of the compound microscope.

An approximate estimate of the magnifying power of the compound miscroscope may be obtained by considering separately the magnification produced by the objective and that produced by the eye-piece. Since the object *XY* on the stage of the microscope is just outside the principal focus of the objective, the image which is at the focus conjugate with this point, is formed far from the lens, towards the upper end of the tube. The size of the image, therefore, is to the size of the object very nearly as the length of the tube is to the focal length of the objective. By the length of the tube is here meant the distance from the stage early to the top of the tube. Accordingly, the magnification due to the objective alone is

$$\frac{x\ y}{X\ Y},$$

or, very nearly,

$$\frac{\text{the length of the tube}}{\text{the focal length of the objective}}.$$

This image is then viewed by the eye-piece as if it were a real object, and, as we have seen above, the magnification produced by this lens is

$$\frac{x'\ y'}{x\ y},$$

or, approximately,

$$\frac{\text{250 millimeters (which is the distance of distinct vision)}}{\text{focal length of the eye-piece}}.$$

From this it follows that the magnification by the two lenses, or, in other words, the magnifying power of the compound microscope, is

$$\frac{x'\ y'}{X\ Y},$$

which is approximately

$$\frac{\text{250 millimeters} \times \text{the length of the tube in millimeters}}{\text{focal length of the objective} \times \text{focal length of the eye-piece}},$$

the focal length of the lenses being expressed likewise in millimeters.

644. The Astronomical Telescope. In the astronomical telescope, Fig. 474, parallel rays of light from the distant object fall upon the object glass *AB* of the telescope, and are brought to a focus at the principal focus. A simple magnifying glass, or eye-piece, *CD*, is then adjusted before the image *xy*, so that the distance of the image from the lens is just a little less than the principal focal length of this lens. A virtual, erect, and magnified image *x′y′* of the image *xy* is produced, and this viewed by by the eye at *E*.

The magnifying power of the telescope may best be obtained by comparing the visual angles which the image and the object subtend at the eye. In determining the visual angle which the object thus subtends, we shall neglect the length of the telescope itself, as being vanishingly small in comparison with the distance of the object. In other words, we shall asume that the visual angle which the object subtends at the object glass is the same as that which it subtends at the eye-piece. The object subtends the angle *XOY*, or θ, and this is the same as the angle *xOy*. Moreover, the image *xy* is situated at the principal focal distance, *Oy*, or *f*, from the objective *AB*. The image subtends the angle *xO′y*, or θ', at the eye, and is situated at the distance *O′y*, which is nearly the principal focal length *f′* of the eye-piece *CD*. *xy* therefore subtends the angle θ in the triangle *xOy*, and the angle θ' in the triangle *xO′y*. By regarding *xy* in each case as an arc subtending these angles respectively, we have (p. 48)

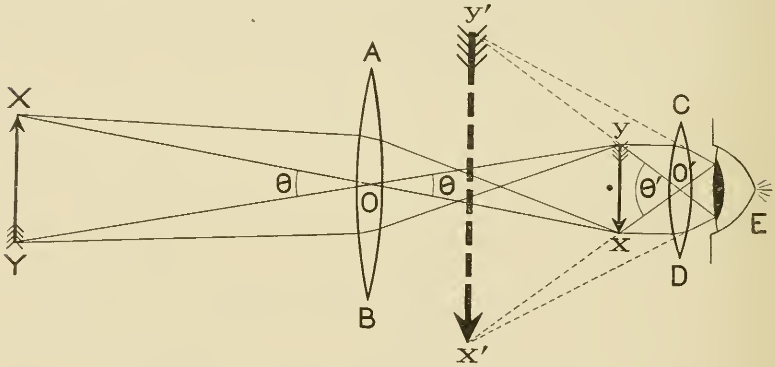

FIG. 474. Diagram of the astronomical telescope.

$$xy = f\theta,$$

and

$$xy = f'\theta'$$

or

$$f\theta = f'\theta'$$

from which

$$\frac{\theta'}{\theta} = \frac{f}{f'} = \frac{\text{focal length of the objective}}{\text{focal length of the eye-piece}}.$$

In other words, the magnifying power of the telescope is practically the ratio of the focal length of the objective to the focal length of the eye-piece.

CHAPTER LV

THE SPECTRUM.

645. The Production of the Spectrum. When a narrow beam of white light passes through a prism and falls upon a screen, the resulting light is not white, but is resolved into a broad band of color, Fig. 475, known as the spectrum. While in reality an infinite number of colors compose this spectrum, and while one shades impreceptibly into the next, there are seven principal colors which may roughly be taken as representing the spectrum, namely, violet, indigo, blue, green, yellow, orange and red. From this we conclude that white light is not homogeneous, but a mixture, being made up of lights of a great many different colors. It is to be observed that the colors occur in the order given above, and that the vio-

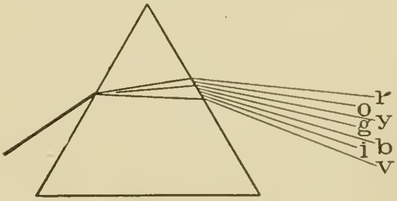

FIG. 475. The refraction and dispersion of white light by a prism.

let is deviated most by the prism, while the red is deviated least. We must infer, therefore, that a prism does not act equally upon lights of different colors, but that it refracts or bends violet light more than it does red. This indicates that in passing through the prism, violet light is retarded more than red light, and consequently is deviated more widely than red from its original direction. The index of refraction of the glass of which the prism is made is thus seen to be slightly different for different colors, being greater for blue light than for red.

This breaking up of a beam of white light into the various colored lights of which it is composed, due to the unequal refrangibility of lights of different colors, is known as *dispersion.* The amount of dispersion depends, therefore, upon the difference between the angles of deviation for violet and for red light. It is more usual, however, to speak of the dispersive power of a prism as the ratio of this difference between the deviations produced for violet and for red light to the mean deviation of the spectrum itself. Thus, if d' represents the angle of deviation of the violet light and d'' this angle for red, while d is the deviation of light from the middle part of the spectrum, then the dispersive power of the prism becomes.

$$\frac{d' - d''}{d}.$$

646. The Recomposition of Light. The fact that white light is composite may be shown by uniting again into one beam the various colored lights produced by the dispersion of white light. If a second prism exactly similar to the first is placed near this one with its refracting edge turned in the opposite direction, Fig. 476, the rays which undergo the greatest deviation by the first prism will likewise be most strongly refracted by the second, and those which are deviated least by the first will be refracted least by the second. Moreover, the bending of the various colored lights in the second prism will be just as great as in the first, and the deviations which are produced are then not only equal, but are also opposite in directon. Accordingly, the second prism recombines the colored lights produced by the dispersion of the first, so that the two together produce simply the effect of a thick plate, and a spot of white light appears upon the screen.

The recomposition of light may likewise be effected by allowing the colored band of light to fall upon a lens. Since the lens brings the various colored lights all to one focus, the different colors are here superposed, and the resulting spot of light is white. In still another way it may be shown that white light is formed by the mixture of several lights. If a disc is painted with the colors of the spectrum and rapidly rotated, the impression formed by each color will persist for some time upon the eye. The eye, therefore, will have the impression of all the colors combined, that is, the various individual colors will disappear entirely, and the disc will then have the appearance of a dull white, or gray.

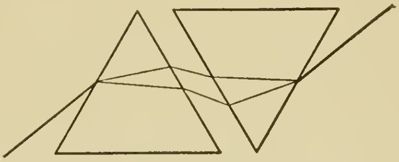

FIG. 476. The recomposition of light by two prisms.

647. The Character of the Spectrum. In order that a spectrum may be observed to the best advantage, care must be taken to have the various colors sufficiently separated so that, as nearly as possible, each may occupy a position by itself in the resulting colored band. Each color must then illuminate only an extremely narrow portion of the screen upon which the spectrum falls, for otherwise an overlapping of different colors would be produced. This may be accomplished by using a prism of high dispersive power, and by providing an exceedingly narrow opening through which light is admitted to the prism, the image of this opening then being focused upon the screen.

Since lights of different colors are refracted by the prism at different angles, it follows that a distinct image of the narrow opening will be produced for each color that is present in the original light. Therefore, if the source of light is one which contains a few colors only, a limited number of images of the opening will be produced in the spectrum. If, however, the narrow opening, or the *slit,* as it is commonly called, is illuminated by white light, which is composed of an infinite number of colors, then an infinite number of lines of light, or images of the slit, are arranged side by side in the spectrum, each overlapping the one adjacent by an amount which decreases in proportion as the slit becomes narrower. Moreover, since the spectrum of white light thus consists of an *infinite* number of partially overlapping images of the source of light, it is obvious that one color must shade imperceptibly into the next, so that the spectrum of white light is *continuous*, that is, it is without vacancies or gaps, from one extremity to the other.

That the spectrum does indeed consist of *an image of the source of light for each color that is present*, may be shown by using openings of different shapes, such as a ring-shaped opening, a V-shaped opening, an S-shaped opening, and so on. If, then, each of these openings is illuminated in turn by red light, or yellow light, or blue light, a distinct image of this opening will be produced in the particular color of the light employed; but if white light instead of colored light is used, a continuous spectrum will be formed, the shape of the spectrum, in each case, being exactly the same as that of the opening.

Accordingly, in the study of the spectrum a narrow slit, which admits but a *line* of light, is used as the source. The various lines of colored light which are then found in the spectrum show the component colors that are present in the original light, and the spectrum formed in this way is said to be *pure*.

648. Color and Wave Length. It has already been suggested, and will be confirmed later, that light is a phenomenon of the vibration of the ether, and that it is propagated through this medium with wave motion. When, therefore, light is dispersed by a prism into its component colors, we are forced to conclude that white light is composed of ether waves of many different lengths, and that certain of these waves travel through the prism more slowly than others, the blue which is refracted most being in this way retarded more than the red which is refracted least. Later it will be shown (p. 869) that the wave length of blue light is only about half as great as that of red, so that this retardation, in general, is greater for the shorter wave lengths. Since the ether waves which constitute the various colored lights are of different lengths, it follows that the vibration frequency of one color is different from that of another, the light in the extreme violet making about twice as many vibrations in a given time as that in the farthest red.

Some idea of the enormous number of vibrations per second made by waves of light may be obtained by estimating the time required of a tuning fork to complete a number of vibrations equally great. If we imagine a tuning fork which makes 1,000 vibrations per second to vibrate continuously day and night, it is found that before the fork would complete as many vibrations all told as break upon the retina of the eye in a single second of time from red light, an interval would have to elapse equal to 12,000 years, while in the same time the blue light makes a number of vibrations twice as great as the red. The matter of color, therefore is purely physiological. The human eye is so constituted that ether vibrations of a certain rate impress our eye with the sensation of red, while vibrations of a more rapid rate affect our eye with a color lying towards the violet end of the spectrum. Just as the ear recognizes two notes of different pitch when these send out different numbers of aerial vibrations, so the eye recognizes two lights of different colors according to the number of ether vibrations which fall upon the retina in a second of time.

649. The Regions of the Ultra-Violet and the Infra-Red. In this connection, the remarkable fact is to be noted that what we actually see of the spectrum does not represent all of the constituent parts of which white light is composed. If a photographic plate is exposed to different portions of the spectrum, it will be found that but little effect will be produced in the red and orange, but that the plate will be acted upon more and more strongly as the violet is approached, while it will be affected most violently entirely beyond the visible spectrum, in the region known as the ultra-violet. Moreover, if a delicate thermometer is carried throughout the spectrum, this instrument will indicate little or no heat in the region towards the violet, but will show a perceptible rise of temperature in the orange and the red. As it is now carried out beyond the visible spectrum into the region known as the infra-red, the temperature will be found to rise, thus showing that, in this direction also, energy in the form of wave motion exists, far out, indeed, beyond the limits within which the human eye is sensitive.

From this we are forced to conclude that our eye can detect but a small portion of the spectrum, and that vibrations exist which are too rapid, as well as those which are too slow, to affect this organ with the sensation of light. The range of the eye, therefore, is really very limited, being confined practically to a visual octave, while it is entirely insensible to ether vibrations of a period either slower or more rapid than those contained within this narrow compass.

It is important, therefore, to remember that the spectrum consists not only of the visible portion, but also of the part lying in the ultra-violet as well as that in the infra-red. Since photographic action is brought about chiefly by the ultra-violet rays, this end of the spectrum is sometimes known as the chemical, or actinic, portion, while the part lying beyond the red, because of the strong thermal radiation there present, is not infrequently called the domain of radiant heat. The larger part of the energy in the spectrum lies in the infra-red, less than one-third in the case of sunlight, and from 3 to 5 percent in the case of gas and incandescent light, being found in the visible portion, while only an infinitesimal amount exists in the ultra-violet. The distribution of this energy, however, is continuous, showing that vibrations of all possible periods are sent out by a source of white light, and therefore that the ultra-violet, the visible, and the infra-red portions of the spectrum are precisely alike in nature, that is, they consist of waves in the ether which differ from one another only as regards wave length. It follows, accordingly, that radiant heat consists of ether vibrations of the same character as those which constitute visible light, the difference between the two being simply one of wave length or period.

650. The Rainbow. By far the grandest and most beautiful example of the refraction and dispersion of light is that which Nature herself furnishes in the rainbow. If the observer stands with his back to the sun and looks upon the falling rain, the magnificient phenomenon observed is the result of the dispersion produced by the refraction of the sun's light in the falling drops of rain. The red of the spectrum is observed to be upon the outside of the bow, and the blue upon the inside, while the center of the bow is on the line drawn from the sun through the observer's eye. Careful measurements show that the angular distance of the bow from this latter line is about 42°.

The theory of the rainbow, and the refraction, reflection, and dispersion of the sun's light within the raindrop may be explained as follows: Let the circle, Fig. 477, represent a section of the raindrop, and let *SI* be a ray of light from the sun falling upon the drop. This light undergoes refraction according to the well-known law; and since blue light is deviated more than red, the white light from the sun is dispersed, the red light passing in the direction *IA*, and the violet light in the direction *IB*. At the points *A* and *B* at the rear of the drop these rays undergo reflection, and pass in the direction *AC* and *BD* respectively. At the points *C* and *D* they are again refracted, and emerge from the drop in the directions *CR* and *DV*. Since in this case the refraction takes place from the denser to the rarer medium, the dispersion which was produced at *I* by the first refraction, is here increased at *C* and *D* by the second.

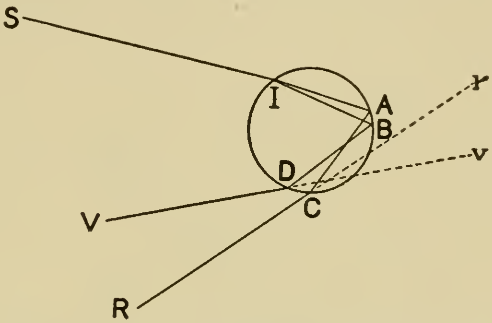

FIG. 477. The refraction, internal reflection, and dispersion of light in a drop of rain, forming the rainbow.

The white light from the sun has therefore undergone two refractions and one internal reflection, and passes out of the drop dispersed, or spread out into the spectrum, of which *V* represents the extreme violet and *R* the extreme red, the remaining spectral colors taking intermediate positions. If the eye were stationed at the point *V*, it would receive the violet *DV*, and the observer would perceive violet light in the direction *Vv*. If, however, the eye were at *R*, it would receive the red ray *CR*, and the observer would perceive red light in the direction *Rr*.

In order, therefore, that the observer may see both the red and the violet rays at the same time, there must be two raindrops, Fig. 478, in each of which this same process of refraction, reflection, and dispersion takes place. Moreover, these two drops must be so situated with reference to each other that the less refracted red ray *C′R′* from the outermost drop will meet at the eye the more strongly refracted violet ray *DV* from the innermost drop. The intermediate colors of the rainbow will then be sent to the eye by drops occupying positions lying between these two. Since the eye always locates an object in the direction of the rays of light which enter it, the eye at *E* sees this spectrum as if it were actually in the position *vr* in the sky.

The refractions and the reflections within the raindrop take place according to perfectly definite laws, and hence the emergent rays make perfectly definite angles with the incident ray. This angle, for the case of the red light of the rainbow, is about 42°. Moreover, since everything is symmetrical about the line *SS″*, drawn from the sun through the eye, and since in every plane through this line raindrops exist bearing to one another and to the eye the same relation as those here described, it follows that, in every plane through the the line *SS″* and at an angle of 42° from it, red light will enter the eye, and at angles slightly less than this other colored lights will also enter the eye.

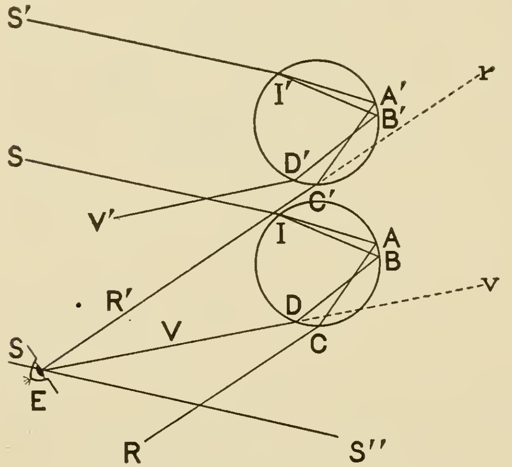

FIG. 478. The relative positions of two drops of rain which send red and violet light to the eye.

In other words, if we imagine the eye as the vertex of a cone described about the line *SS″* as an axis, the elements of the cone making an angle of 42° with this line, Fig. 479, the drops of rain which at any instant are found upon the surface of this cone will disperse the white light from the sun and will send red light to the eye. Moreover, upon the surface of a cone of slightly smaller angle, drops of rain will be found which bear the same relation to the drops on the outer cone that the lower drop, shown in Fig. 478, bears to the upper one, and these drops will then send violet light to the eye. Since we always perceive an object in the direction of the rays of light which enter the eye, the eye projects these rays against the background of the sky; and as the drops which thus send these rays to the eye are found upon the surface of a cone, and as the projection of this conical surface is a circle, the band of color which is then seen in the sky is circular in form, and the rainbow is therefore circular.

Furthermore, since these particular drops send the rays here designated to the eye of one observer, other drops upon the surface of another cone will send corresponding rays to the eye of another observer, so that the two observers will see different bows. The rainbow, therefore, is not a fixed object in the sky to be observed as a material thing, but is a subjective phenomenon which must be formed by each eye for itself, inasmuch as each eye that sees a rainbow must be at the vertex of a cone, upon the surface of which lie the particular drops of rain that form this bow.

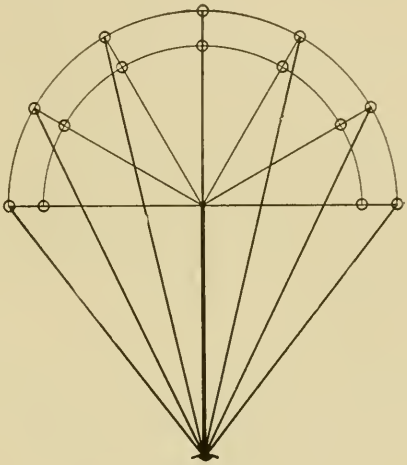

FIG. 479. Diagram explaining the circular form of the rainbow.

The description given above applies only to the primary bow. It is often possible, however, to see, outside of the primary bow, also a fainter, or secondary, bow, with colors in the reverse order, and at an angle of 51° from the line drawn from the sun through the eye. This bow is produced by rays from the sun which are incident upon the drop at a different point from those which give rise to the primary bow, and which undergo two refractions and two internal reflections within the drop. The path of the rays of light forming the secondary bow may also be readily followed, Fig. 480. The ray of sunlight, *SI*, which is incident upon the drop at *I*, is refracted, the red light *IA* being deviated less than the violet light *IB*. At *A* and *B* these rays are reflected upon the smooth interior surface of the drop, and at *C* and *D* they again undergo reflection. At *E* and *F* they are refracted a second time, and pass out of the drop in the direction *ER* and *FV*, the original dispersion produced at *I* being increased by the second refraction at *E* and *F*.

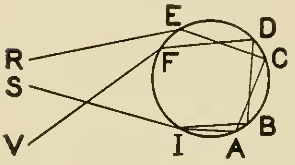

FIG. 480. The formation of the secondary bow.

In the formation of the secondary bow, the light undergoes one more internal reflection than in the primary bow, and consequently the colors are fainter. Morever, since the light is incident at the lower part of the drop in the secondary bow, and in the upper part of the drop in the primary bow, the refraction in the two cases takes place in opposite directions, hence the colors in the secondary bow are in the reverse order of those in the primary bow, as may be seen at once by an examination of the diagram.

651. Dispersion Without Deviation—The Direct-Vision Prism. In discussing the separation of the various colors produced by the refraction of light, it is necessary to answer the question as to whether dispersion is dependent upon refraction and proportional to it, or whether the dispersion of a prism depends upon the material of which the refracting substance is made.

This question may be answered experimentally by providing three prisms, one of flint glass, one of crown glass, and a hollow prism filled with water, all having equal refracting angles. Upon sending a ray of light through each of these prisms, it will be seen that the flint glass prism produces the greatest deviation of the light, the crown glass prism nearly but not quite so much, and the water prism least of all. Furthermore, it will be seen that the spectrum produced by the flint glass prism is about twice as long as that developed by the crown glass prism, while the latter is much longer, than the one yielded by the water prism. From this it is obvious that the dispersion is not proportional to the deviation of the light, but that it does depend upon the material of which the refracting substance is made.

This disproportionality, which always exists between refraction and dispersion, is known as the *irrationality of dispersion.*

In order that a flint glass prism may produce the *same angular deviation* for the middle of the spectrum (the yellow light) as a crown glass prism of 60°, the flint glass prism must be only 52°. While, therefore, the actual deviation for the middle of the spectrum produced by a flint glass prism of 52° is equal to that formed by a crown glass prism of 60°, the length of the spectrm produced by the flint glass prism is nearly double that of the spectrum developed by the crown glass prism. Accordingly, if a flint glass prism of 52° and a crown glass prism of 60° are placed together with their refracting edges turned in opposite directions, Fig. 481, the bending of the yellow light produced by the flint glass prism in one direction will be just corrected by the deviation produced by the crown glass prism in the opposite direction, and, as a result, the light in the middle of the spectrum (the yellow light) will pass through undeviated.

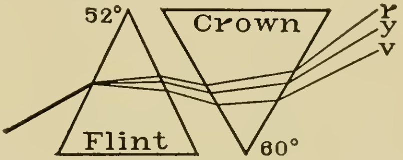

Fig. 481. Diagram of the direct-vision prism.

While, however, the deviation is thus compensated, *the color is not removed,* since the spectrum produced by the flint glass prism is nearly twice as long as that yielded by the crown glass prism. In other words, the dispersion of the 52° flint glass prism is only about half removed by the feebler dispersion in the opposite direction of the 60° crown glass prism, and therefore a spectrum will be formed the length of which is the difference between the lengths of these two. Moreover, the colors will appear in the order in which they would be formed by the flint glass prism alone, so that it is the flint glass prism which determines the arrangement of the colors in the resulting spectrum.

By thus combining two prisms of unequal refractive and dispersive powers, it is possible to compensate for the angular deviation produced by refraction, and therefore to bring about *dispersion without deviation.* The combination is often of the greatest service, and forms what is known as the *direct-vision prism.*

652. Deviation without Dispersion—Achromatism. Because of the irrationality of dispersion, it is possible, also, by combining two prisms of different refracting angles and different materials, to produce a deviation of the ray without dispersion. Experiment proves that a flint glass prism of only 30° will produce a spectrum of the same length as a crown glass prism of 60°, although the deviation will be much greater in the latter case than in the former. Therefore, if a flint glass prism of 30° and a crown glass prism of 60° are placed with their refracting edges turned in opposite directions, Fig. 482, the dispersion which the flint glass prism causes is just removed by the equal dispersion in the opposite direction which the crown glass prism produces, and the resulting beam of light is colorless. The actual bending of the light, however, brought about by the 30° flint glass prism, is less than that produced in the opposite direction by the larger-angled crown glass prism, and hence the resulting beam of light, although colorless, is deviated, the bending taking place towards the thicker portion of the crown glass prism. This deviation is therefore less than that which would have been produced by the crown glass prism alone, and indeed, is equal to the difference between the deviation caused by the crown glass prism and the slighter deviation in the opposite direction produced by the smaller-angled flint glass prism. Such a combination of two prisms, producing *a deviation of the ray without dispersion,* is known as the *achromatic prism.*

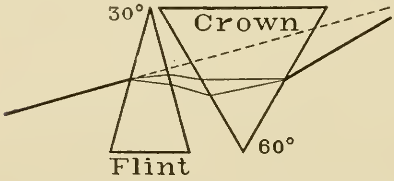

FIG. 482. Diagram of the achromatic prism

653. Chromatic Aberration. From the fact that the angle of deviation for violet light is greater than that for red, it is evident that when white light passes through a single lens, the violet rays will be brought to a focus nearer the lens than will those from the red end of the spectrum, while lights of intermediate colors will be brought to foci at intermediate positions. Therefore, on account of the unequal refrangibility of the different colored lights, a single lens cannot give a colorless, and therefore a distinct, image, inasmuch as lights of different colors are brought to foci at different distances from the lens, Fig. 483.

If the screen upon which the image falls is placed at V, which is the focus for the violet light, there will be seen a bluish image surrounded by an indistinct red border. So, also, if the screen is placed at R, the focus for the red rays, there will be formed a reddish center surrounded by a blue border. This coloration of the image and the confusion and indistinctness thus produced, due to the fact that the differently colored rays come to foci at different distances from the lens, is known as *chromatic aberration*. It is obvious that chromatic aberration cannot be avoided by the use of a single lens.

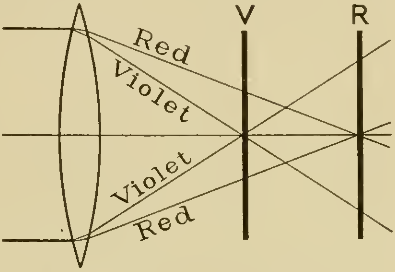

FIG. 483. The cause of the chromatic aberration produced by a single lens.

654. The Achromatic Lens. It is possible, however, by combining two lenses of different refractive and also of different dispersive powers, so completely to remove the dispersion, that the resulting image will be almost colorless. Since the dispersion of flint glass is nearly twice as great as that of crown glass for the same amount of refraction, a crown glass lens of a certain focal length and a flint glass lens of a curvature half as great, and therefore of twice this focal length, would produce equal amounts of dispersion, or equal chromatic aberration. Therefore, if a convex crown glass lens of a certain focal length, Fig. 484. is combined with a concave flint glass lens of twice this focal length, the chromatic aberration will be removed, while the focal length of the combination will be greater than the focal length of the crown glass lens alone. Such a combination of lenses which produces an image practically free from color is known as an *achromatic* lens.

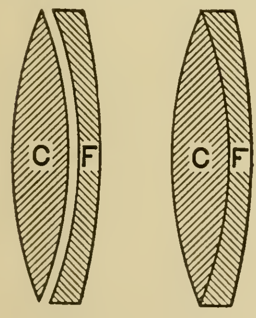

FIG. 484. The achromatic lens.

The principle of combining two lenses to remove the chromatic aberration which is inevitable in the use of a single lens, was first discovered by Hall in 1733. This great discovery, however, seems to have remained unnoticed for a number of years, and the principle was rediscovered in 1757 by Dolland, who is commonly regarded as the originator of the achromatic lens. The ability thus to produce an image free from color has made possible research and discovery in all branches of science.

655. The Different Types of Spectra. An examination of spectra shows that all may be divided into three classes: namely, continuous spectra, discontinuous, or emission, or bright-line, spectra, and dark-line, or reserved, or absorption, spectra. These three classes will be studied in order.

656. Continuous Spectra. When white light from a narrow slit passes through a prism and falls upon a screen, we have seen that it is dispersed by the prism into a band of color called the spectrum. We have observed, moreover, that this spectrum is perfectly continuous from end to end, since every possible wave length exists in the white light, and each wave length casts upon the screen an image of the slit in the color corresponding to that wave length. The spectrum produced by white light is therefore perfectly continuous, since all possible wave lengths are present. This, then, is the spectrum emitted by *a solid or a liquid* which is heated to incandescence, and may be seen to the best advantage by dispersing the light emitted from the white hot crater of the electric arc.

657. The Cause of the Continuous Spectrum. When the properties of a heated body are investigated, the remarkable fact is developed that all substances when incandescent, that is, when raised to a temperature at which they become self-luminous, are found to be conductors of electricity. This can be interpreted only by assuming that *free electrons,* which then act as carriers of electricity, *exist among the particles of the heated body.* We believe, therefore, that, as we raise the temperature of a body, the increased molecular bombardments which take place between the molecules disengage or set free certain electrons formerly existing within the atoms, so that these electrons vibrate back and forth among the molecules of the body, thereby rendering the latter a conductor. Moreover, these electrons, by their collisions and consequent changes of velocity, are continually accelerated, and therefore give rise to ether waves which go out from these particles as centers. Since these electrons are extremely irregular in their collisions, and therefore in their accelerations, they send out ether waves of all possible periods, so that these waves, when analyzed by a prism, give rise to a spectrum containing all possible wave lengths.

If the heated body is not yet incandescent, the ether waves of the periods which are able to produce the sensation of light are still too feeble in energy to affect the eye, and only those of longer wave lengths lying in the infra red, where the largest amount of energy is found, can be detected. As the body is now raised in temperature until it begins to be self-luminous, greater numbers of electrons are released, and these, by their bombardments, increase the energy, that is, the amplitude, of the ether waves which are thus emitted. The energy of those waves which affect the eye with the sensation of red is then sufficiently great to be detected, although the waves towards the blue end of the spectrum are still too feeble to be perceived, and the body is then said to be red hot. At a still higher temperature the waves farther towards the blue likewise increase in energy and become successively visible, until, when the temperature of full incandescence is reached, the waves even in the extreme violet possess an amount of energy which enables them to be seen, and the body is then white hot.

Moreover, this effect, which is accomplished by the wholly irregular acceleration of the free electrons, can be attained only when the electrons thus bombard confused masses of molecules, that is, when the body which emits this light is either an incandescent solid or liquid. Furthermore, when the complex assemblage of all the vibrations which make up white light undergoes refraction by the prism, the various colors corresponding to these ether waves are arranged side by side in the order of their wave lengths; and inasmuch as all possible wave lengths are present, no gaps or vacancies occur in the colored band, and the spectrum is continuous. It follows, therefore, that *a continuous spectrum is produced by an incandescent solid or liquid.*

658. The Bright-Line, or Emission, or Discontinuous, Spectrum. When the light from an *incandescent vapor or gas* is analyzed by a prism, the spectrum is discovered to be *not continuous, but to consist of a number of bright lines,* separated from one another by dark spaces. Moreover, these bright lines are found to differ in number and order for each incandescent gas examined, although the particular lines for any one gas are always found in the same position in the spectrum.

Thus, if a metal is volatilized in the non-luminous flame of the Bunsen burner or in the electric arc, the vapor of this metal, which thus shines by its own light, does not emit light of all wave lengths, but sends to the screen a few wave lengths only, and as a result but a limited number of bright lines appear upon the screen, one line for each different wave length present in the light emitted. *This assemblage of bright lines separated by dark spaces is called the bright-line, or the emission, or the discontinuous, spectrum, and is produced whenever light from a glowing gas or vapor is dispersed by a prism.* The Thomson atom (p. 649) not only lends itself to an explanation of this phenomenon, but serves also to remove many of the difficulties which had previously been encountered in the electromagnetic theory of light.

659. The Cause of the Bright-Line, or Emission, Spectrum. After a solid or liquid has been raised to incandescence, the further addition of heat so increases the kinetic energy of the particles that the inter-molecular bonds are broken, and the body is volatilized, that is, it passes from the state of the solid or liquid into the condition of an incandescent gas. Moreover, at this exceedingly high temperature, the molecules themselves can no longer remain intact, since the atomic bonds, which join together the atoms to form the molecules, are then likewise ruptured, and the gas is said to dissociate. In an incandescent dissociated gas, therefore, the atoms exist as free vibrating particles, and bombard one another with an energy which is proportional to the temperature to which they are raised.

So violent, indeed, are these inter-atomic collisions, that they are believed to affect the rotation of the electrons within the atoms. This may be accomplished either by producing an unequal distribution of the electrons which revolve in a given orbit, or causing them, in their revolution, to deviate from the plane of their orbit, that is, to revolve with an undulatory movement, rising alternately above and below their otherwise uniform plane of motion. In either case, the electrons change their velocity or their direction of motion, or perhaps both, and therefore are accelerated, in consequence of which a radiation in the form of ether waves goes out from these electrons as centers.

Moreover, the accelerations which the electrons undergo depend upon the number and arrangement of these particles within the atom, that is, upon the nature and constitution of the atom. It follows, therefore, that an atom of one substance will send out a particular radiation which is peculiar to that substance, while an atom of another substance will emit a radiation which is altogether different. Furthermore, since these disturbed electrons are limited in number and are periodic in their acceleration, they give rise to a perfectly definite number of regular vibrations in the adjacent ether, a wave of a particular length being formed for each accelerated electron or group of electrons present in the atom. Accordingly, the radiation which is thus emitted does not consist of all possible vibrations, but of a few wave lengths only, and these, when this light is analyzed by a prism, then fall at perfectly definite places in the spectrum.

In view of the fact that a spectrum consists of a series of images of the slit, that is, a line of light for each wave length present, it is seen that the spectrum of a glowing gas or vapor cannot be continuous, or, in other words, cannot consist of all wave lengths, since only a few wave lengths are sent out by the source. Each ether vibration, however, produces a bright line of light at a position in the spectrum depending upon its wave length, while totally dark spaces occur between these bright lines, inasmuch as the wave lengths which would produce light in the latter regions do not exist. Accordingly, the spectrum of the light emitted by an incandescent gas consists of a series of bright lines separated by dark spaces, the number and color and arrangement of these lines depending upon the character of the particular substance in the gaseous condition which is emitting the light.

Thus, for instance, when hydrogen is rendered self-luminous by the electric discharge, the light which it then emits is found to contain mainly three colors or wave lengths, and the spectrum formed by this light consists of a bright line in the red, a bright line in the region between the green and the blue, and a somewhat weaker line in the violet. The ocurrence of these lines, which always appear at precisely the same positions in the spectrum, indicates the presence of glowing hydrogen.

So, also, if the metal sodium is volatilized in the flame of a Bunsen burner, the incandescent sodium vapor emits light of practically one wave length only, and hence but a single bright line, and that in the yellow, is produced in the spectrum of sodium. In further resolution, this single yellow line is, indeed, found to be double, and therefore to consist of two lines lying very near together. At the temperature of the electric arc, the electrons in the sodium atom are thrown into a more complicated vibration, and hence at this temperature there appear other lines in the sodium spectrum in addition to the principal yellow line.

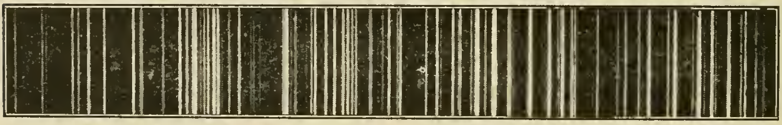

FIG. 485. A small portion of the bright-line, or emission, spectrum of iron.

In the same way, the vapor of copper emits light chiefly of one wave length, and this gives a single line in the green. The electrons in the silver atom, when this metal is in the state of vapor, are found to have two principal periods of vibration, and as a result the spectrum of silver consists of two beautiful lines, one in the green, and the other between the green and the blue.

660. The Zeeman Effect. A much clearer interpretation of the phenomenon of the bright-line spectrum, and at the same time a confirmation of the existence of an atom analogous, at least, to the one conceived by Thomson, is given by the remarkable discovery, made by Zeeman in the year 1896, that the character of the light emitted by an incandescent gas is to some extent modified by a magnetic field. Zeeman showed that when an incandescent gas is subjected to the action of an intense magnetic field, certain lines in the spectrum of this gas are doubled, trebled, or even quadrupled, depending upon the intensity and direction of the field. This evidently points to the conclusion that the vibrating particle generates a magnetic field in the ether about it, that is, that this particle is highly electrified and in rapid motion, so that it produces the effect of an electric current. Moreover, the direction of the displacement of the lines in the Zeeman effect would point to the fact that the vibrating paricle is negatively electrified.

The entire phenomenon therefore gives strong experimental evidence of the existence of an atom which has many of the properties of the Thompson atom, and practically establishes the electromagnetic theory of light, which was predicted mathematically by Maxwell. From every point of view, therefore, it it highly probable that the vibrating particles which give rise to the bright lines in the spectrum *are the electrons themselves, as they vibrate according to their own law within the atom.* Moreover, this explains the fact that the spectrum of a substance is perfectly characteristic of that substance, for we should then assume that the attributes and special properties of an atom are determined by the arrangement of the electrons which the atom contains and the law according to which these electrons vibrate. That this law, in some instances, must be extremely complex may be inferred from the many hundreds of lines which appear in the spectrum of a single substance, such as iron, Fig. 485.

661. The Dark-Line, or Reversed, or Absorption, Spectrum. The dark-line, or reversed, or absorption, spectrum, of which the solar spectrum is the great type, is one that is otherwise continuous, but crossed by large numbers of fine dark lines, Fig. 486.

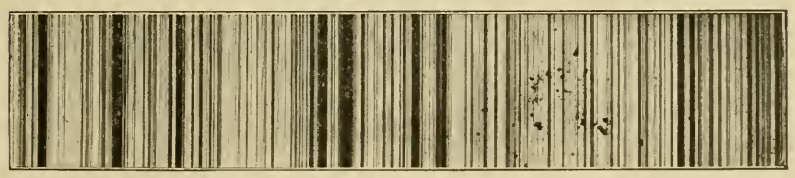

FIG. 486. A portion of the solar spectrum, showing the Fraunhofer lines.

When the light from the sun is analyzed by a prism, it is found to consist of a continuous spectrum, with, however, a large number of dark lines superposed upon it. These lines were first discovered by Wollaston in the early part of the nineteenth century, and studied much more in detail later by Fraunhofer. The latter investigator found that these dark lines, which cross the bright and otherwise continuous spectrum of the sun, always occur in precisely the same places in the solar spectrum, and he gave the names of the letters of the alphabet to a number of the most prominent, beginning with the line in the deepest red. The letters which he assigned to these lines have always been retained, and these dark lines in the solar spectrum are still known as the Fraunhofer lines. Fraunhofer could offer no explanation for the existence of these lines, and it was not until the work of Foucault, in 1849, and the researches of Kirchoff and Bunsen, in 1859, that the mystery of the dark lines in the solar spectrum was explained.

The coincidence between many dark lines in the spectrum of the sun and numerous bright lines in gaseous spectra had long been observed. Kirchoff and Bunsen showed that the dark lines occur rigidly at the same parts of the spectrum at which bright lines are produced when lights from certain incandescent vapors are analyzed by a prism. By placing a glass tube containing sodium so that the white light from the flame of a candle should pass through it, these investigators showed that the white candle light would produce a continuous spectrum so long as the tube was cold. Upon heating the tube, however, so that the latter was filled with the vapor of sodium, they observed that two fine dark lines appeared at exactly the same points in the spectrum at which the two bright lines of sodium would appear when light from incandescent sodium vapor was examined by itself. This showed that the sodium flame allows light of all wave lengths to pass through it except light of the two wave lengths which it itself emits, and these it absorbs. In other words, the sodium flame is transparent to light of all other wave lengths except the particular ones which it emits, but is opaque to light of its own wave lengths.

This phenomenon, however, is simply in accord with the very general principle, that a vibrating particle of any definite period will absorb all other vibrations of a period exactly equal to its own, as may be seen when one pendulum in vibration takes up the motion of another of the same period, or when one tuning fork vibrates symphathetically with another, and absorbs its energy. Therefore, when white light from an incandescent solid passes through a *cooler* column of incandescent sodium vapor, this vapor will absorb from the white light those vibrations which have periods equal to those in which it itself vibrates. The light, therefore, which reaches the screen is white light minus light of the particular wave lengths which sodium emits, since these were absorbed from the white light as the latter passed through the sodium vapor. Accordingly, the resulting spectrum will be continuous with the exception of two dark lines in the yellow at exactly those points at which bright yellow lines would appear if the sodium alone were burned in a flame without the white light behind it.

Careful observation, however, will show that these dark lines are not black but appear so only by comparison with the much brighter light in the adjacent portions of the spectrum. Indeed, these dark lines are really bright lines, since they are the light which is sent to the screen by the incandescent sodium vapor; but since the latter is cooler than the source of white light, the light which it emits is less intense than that which comes from the more highly heated source behind it. If the sodium vapor were at the same temperature as the source of white light, it would send to the screen as much light as it absorbs, and the dark lines would disappear. *The condition, therefore, for the production of the dark-lines in the reversed, or absorption, spectrum is that white light from an intensely heated source must pass through a column of gas or vapor of lower temperature.* This column of cooler gas is, indeed, slightly raised in temperature by the energy which it thus absorbs from the white light; but as it sends out radiation in all directions, the additional amount which it emits in the particular direction of the screen is too slight to be detected, and the dark absorption lines remain practically of the same intensity as that with which they would appear if the white light were absent.

662. The Cause of the Fraunhofer Lines. The discovery of Kirchoff and Bunsen offered at once a ready and complete explanation of the existence of the dark lines in the solar spectrum. It was then seen that the white light from the incandescent solid or liquid body of the sun must pass through clouds of vapors of many of the elements in its passage to the earth. It is therefore apparent that the intensely white surface of the sun, called the photosphere, must be surrounded by clouds of metals in the state of vapor, due to the extreme heat. This layer, which is called the chromosphere, absorbs, from the white light emitted by the photosphere, light of the particular wave lengths which these elements themselves would emit; and therefore the light from the sun comes to us minus light of a great many wave lengths, and for each wave length which has been thus absorbed, a dark line is produced in the spectrum. This occurrence of the dark solar lines at the very places at which bright metallic lines appear when the various terrestrial elements are volatilized, enables us to study with great precision the composition of these clouds of vapor which surround the sun. Thus, Fig. 487 is from a photograph of the dark-line solar spectrum, across the center of which has been photographed at the same time the bright-line emission spectrum of iron. The exact coincidence of these two series of lines is conclusive proof that iron in the state of vapor is one of the many elements which enter into the composition of the solar atmosphere. In a similar way it has been shown that not a few of the Fraunhofer lines are due to the absorption of light of certain wave lengths, as the light from the sun passes through the various gases and the water vapor of our own atmosphere.

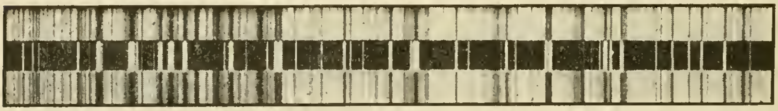

FIG. 487. A portion of the bright-line spectrum of iron photographed upon the dark-line solar spectrum, showing the presence of iron in the clouds of vapor which surround the sun.

Many of the stars likewise have dark-line, or absorption, spectra, while a few show bright-line, or emission, spectra, the latter being also the type of spectra emitted by the nebulae. From a comparison of these lines with those occurring in the same positions in the spectra from known sources, it is possible to obtain not only a definite knowledge of the chemical composition of the celestial bodies, but also to some extent an idea of their physical condition as well as their relative age and development.

663. Observations of the Protuberances upon the Sun. For many years, during the times of total eclipses of the sun, rose-colored protuberances, resembling great sheets and tongues of flame, have been observed to exist around the edge of the sun, when the more intense light from the solar surface was intercepted by the dark body of the moon. Although these red flames or protuberances extend 3,000 or 4,000 miles above the body of the sun, they cannot ordinarily be seen, owing to the relative faintness of the light which they emit. During the total eclipse of 1868, these protuberances were examined with the spectroscope and were found to consist of sheets and tongues of incandescent vapor, in which the bright hydrogen lines were noticeably prominent, thus showing that these prominences consist, in part at least, of flames and clouds of glowing hydrogen. Immediately after the eclipse, it occurred simultaneously and independently to Janssen and Lockyer that with the aid of the spectroscope these protuberances could be observed any clear day, even when the sun was not eclipsed. Experiment showed that this is indeed possible; and since that time observatories have kept a daily record of these red flames as they are formed at irregular periods of time around the edge of the sun.

The observation of the prominences upon the uneclipsed sun depends upon the principle that the intensity of the light in any part of a continuous spectrum is weakened by increased dispersion. When white light is analyzed by a single prism, it is spread out into a continuous spectrum of limited length, and therefore of a fair degree of brightness. When, however, white light is passed through a great number of prisms, the length of the spectrum is correspondingly increased, due to the increased dispersion, while the brightness of any part is decreased in the same ratio. On the other hand, the width of a bright line in a bright-line spectrum is not sensibly altered by higher dispersion, since this line contains light of but a single wave length, and therefore retains its original brightness, even though the light may be passed through a train consisting of several prisms.

In order, therefore, to observe the solar prominences in direct sunlight, the image of the edge of the sun is so thrown upon the slit of a spectroscope containing many prisms that part of the slit receives light from the outermost edge of the sun, and part is illuminated by the light from the prominences themselves. The white light from the edge of the sun and also that produced by the illumination of the earth's atmosphere, as well as that which is emitted by the solar corona, are all weakened by the very high dispersion employed, while the light from a protuberance, consisting, as it does, principally of light from glowing hydrogen, gives simply the bright-line spectrum of hydrogen, and this is not sensibly weakened by its passage through the large number of prisms. It follows, therefore, that if the slit of the spectroscope is opened sufficiently wide to receive the image of the prominence, *this image itself and not the slit now becomes the source of light,* and this will produce an image of the protuberance at positions in the spectrum occupied by each of the three wave lengths which are emitted by glowing hydrogen gas. Accordingly, there will be seen in the spectroscope an image of the prominence in the red of the spectrum, one in the region between the green and the blue, and one in the violet. These bright-line images of the protuberance are all superposed upon the very faint continuous spectrum of the light reflected from the solor corona and from the earth's atmosphere, while on the side of the prominence towards the sun there will also be seen the faint continuous spectrum of the edge of the sun, all of these continuous spectra being greatly weakened in intensity by the very high dispersion employed.

The clouds of vapor about the sun which send out these tongues of flame are the clouds of the chromosphere, and it is in these clouds that light of certain wave lengths is absorbed from the white light emitted by the white-hot body of the sun. This layer of clouds to which the prominences belong is therefore sometimes called the reversing layer around the sun.

664. Motion in the Line of Sight—The Doppler Effect. If a source of light is fixed in position, and an observer who perceives this light is also stationary, a certain number of ether waves will be received each second, and a given wave length of light, when analyzed by a prism, will produce a line in the spectrum occupying a perfectly definite position. If, however, the source of light and the observer move towards each other, a greater number of vibrations per second will meet the eye, and therefore the line in the spectrum corresponding to this light will be shifted in the direction of the more rapidly vibrating light, that is, towards the violet end of the spectrum. On the other hand, if the source of light moves away from the observer, not so many vibrations per second will be received, and therefore the line in the spectrum which this light produces is displaced in the direction of light making a smaller number of vibrations per second, that is, toward the red end of the spectrum. This phenomenon is in accordance with the principle discussed under the subject of sound (p. 690), and there illustrated by corks being thrown into a stream. If the person throwing in cork moves down stream, the corks will come more frequently to the observer stationed below, and if he moves up stream they will come less frequently, than if he were to remain stationary.

The light which comes to us from the stars is found to give spectra which may, in general, be divided into two classes, namely, bright-line spectra and dark-line spectra. Thus, the light from the bright star Sirius gives a continuous spectrum crossed, however, by dark lines, among which may be recognized three as occurring in the positions occupied by the hydrogen lines. This indicates that in all probability the star Sirius is an incandescent solid or liquid body, surrounded by a dense atmosphere containing glowing hydrogen gas. Careful observation shows, however, that these dark lines are not exactly in the positions occupied by the hydrogen lines, but that they are slightly displaced towards the red end of the spectrum. The star and the earth, therefore, are moving "in the line of sight" with respect to each other; and from the displacement of the lines towards the red end of the spectrum, it is evident that the star is receding from us. Moreover, from the amount by which these lines are displaced, it is known that the velocity of recession is about 20 miles per second. In the same way, the lines in the spectrum of the star Arcturus are found to be displaced by a considerable amount towards the blue, thus showing that this star is approaching the earth at the rate of about 50 miles per second.

In a very general way, it is found that the spectral lines of all the stars in one part of the heavens are displaced towards the blue end of the spectrum, while those of the stars in the opposite part of the skies are displaced by equal amounts towards the red. From this it may be concluded that the entire solar system is not stationary in space, but is sweeping on towards a certain point among the stars with an enormous velocity.

Döppler conceived the idea of explaining the different colors of the stars by their relative motions in the line of sight with respect to the earth, assuming that the bluish stars are those which are moving towards us, and the reddish stars are those which are moving away from us. While, indeed, it is true that the apparent wave lengths of the light, and therefore also, to some extent, the colors of the stars, are affected by their motion in the line of sight, the change in the color thereby produced is so small as to be wholly without influence upon the appearance of the stars as we see them without spectroscopic aid.

665. Color. Objects which are not self-luminous are seen by the light which they irregularly reflect. Those bodies which we ordinarily call white are able to reflect all colors nearly in the same proportion, and, accordingly, if they are illuminated by white light, they appear white, since light of all color is reflected to the eye. On the other hand, if a white body is illuminated by monochromatic light, the body will appear of the same color as the light itself. In general, however, bodies have the power of reflecting certain colors more copiously than others, and hence these bodies appear colored when illuminated by white light, the color of the body being a mixture of all the colored lights which that body reflects, while those colors which are not reflected are absorbed. A gray body is one which reflects only a small part of each color of the white light that falls upon it and absorbs the rest, hence gray may be regarded as a very dull white. A black body is one which absorbs all the radiations which fall upon it and reflects none, either in whole or in part. White, gray, and black, therefore, may be regarded as all being of the same color, and differing from one another only in brightness. If a red ribbon is held in the various colors throughout the spectrum, the ribbon will appear black everywhere except in the red, since the red ribbon has the power of reflecting red light only, all other colored lights being absorbed.

When white light falls upon a transparent body, such as a piece of colored glass, certain colors are transmitted, while the remainder are absorbed. The color of a transparent body, therefore, is the result of the mixture of those colors which pass through the body. A colored transparent body does not color the light which traverses it, that is, it does not add anything to the white light, but simply acts as a filter to remove from the white light all the other colors except the ones which it transmits. All opaque powders or pigments are made of particles which are really transparent, as may be proved by observing thin slices of the materials of which these colored pigments are made before they are ground to powder. The light which a colored pigment sends to the eye is therefore a mixture of those colors which have penetrated to a certain depth beneath the surface of the pigment, and then been reflected to the eye.

When light falls perpendicularly upon the surface of uncolored glass, only about one-twenty-fifth part is reflected. If a second surface exists beneath the first, about one-thirteenth will be reflected, and if the number of reflecting surfaces is very great, nearly all of the light will be reflected. This must also be true when the glass is ground to a white powder. Therefore, of the white light which is reflected from the white powder, only a small part comes from the first surface, the remainder being reflected from the surfaces of particles at greater depths. It is thus seen that if the powder is colored, that is, if it has a selective absorption, certain colors will be removed in their passage through the particles of the powder or pigment, and hence only those colors which are not thus absorbed will be reflected from the interior layers. It is, therefore, the mixture of these reflected colors which produces the color of the pigment.

666. The Mixture of Colored Pigments. If light were to pass in succession through two differently colored transparent bodies, such as two pieces of colored glass, each would possess a selective absorption of its own, and hence only those colors which are common to the two would be transmitted. Thus, a piece of what is ordinarily called yellow glass will transmit colors in the red end of the spectrum, that is, the red, orange, yellow, and green, and will absorb the violet, indigo, and blue, while a piece of blue glass will transmit colors in the blue end of the spectrum, namely, the violet, indigo, blue, and green, and will absorb the red, orange, and yellow. Accordingly, if light is made to pass first through the yellow glass, only the red, orange, yellow, and green will be transmitted, since the violet, indigo, and blue are absorbed. If this transmitted light is now made to fall upon a piece of blue glass, the latter absorbs the red, orange, and yellow, and only the green can pass. Therefore, the combination of yellow and blue glass will transmit only green light.

It is upon this principle that the formation of colors produced by the mixture of colored pigments depends. A yellow pigment is one that absorbs from the white light falling upon it all the blue end of the spectrum and reflects only the red, orange, yellow, and green, while a blue pigment is one that absorbs from white light all the red, orange, and yellow, and reflects only the violet, indigo, blue, and green. Accordingly, if a blue and a yellow pigment are mingled, the yellow pigment will absorb from the white light all the violet, indigo, and blue, while the blue pigment will absorb all the red, orange and yellow, and hence the green, which neither pigment absorbs, and which, therefore, they both reflect, will be the color of the mixture. This, may, perhaps, be illustrated as follows: If v, i, b, g, y, o, r, indicates the white light which falls upon a mixture of a blue and a yellow pigment, the light reflected from the blue pigment may be represented by v, i, b, g, —, —, —, while, in a similar way, the light reflected from the yellow pigment may be denoted by —, —, —, g, y, o, r, where — indicates a color which the pigment absorbs.

Therefore,

Light reflected from the blue pigment = v, i, b, g, —, —, —.
Light reflected from the yellow pigment = —, —, —, g, y, o, r.
Light reflected in common = g,

Accordingly, *the color produced by a mixture of pigments* is seen to be the difference between the colored lights which the two pigments reflect, that is, *it is the color which neither absorbs, and therefore the one which the two reflect in common.*

667. The Mixture of Colored Lights. The effect, however, produced by mixing colored lights is entirely different. Thus, if a beam of white light is passed through a piece of yellow glass, light containing the colors, red, orange, yellow and green will fall upon the screen; and if another beam of white light is passed through a piece of blue glass, light of the colors violet, indigo, blue and green will fall at another place upon the screen, Fig. 488. If the yellow and the blue circles of light are now made to overlap, that part of the screen which contains the superposed lights will send to the eye red, orange, yellow and green light from the yellow beam, and violet, indigo, blue and green light from the blue beam, so that the eye receives light of all colors from that portion of the screen where the two beams overlap, and the effect is therefore that of white light. This may be shown diagrammatically as follows:

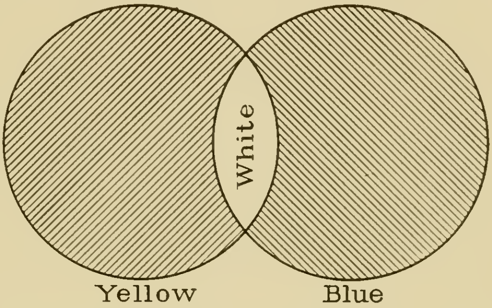

FIG. 488. The mixture of yellow light and blue light produces white light. Yellow and blue are therefore complementary colors.

Colors received from blue light = v, i, b, g,
Colors received from yellow light = g, y, o, r.
Colors received from both lights = v, i, b, g, y, o, r (white light)

The resulting color produced by a mixture of colored lights is thus seen to be the sum of the colors which each light would produce alone.

The colors of those lights which, when mixed together, produce white light are called *complementary colors.*

CHAPTER LVI

THE INTERFERENCE OF LIGHT

668. The Production of Interference. In our study of the propagation of waves, we have seen that if two vibratory motions affect the same particle at the same time, the particle must suffer a displacement which is the algebraic sum of the component displacements. If the two vibrations are of the same period and in the same phase, the resulting amplitude is greater than that of either alone; but if the vibrations are of the same period and amplitude and opposite in phase, the algebraic sum of these motions is zero, and the particle does not move. This combination of two vibratory motions to produce a vibration of increased or of decreased amplitude is known as interference. In the study of sound, also, we have seen that two sound waves may be made to combine so as to produce a sound louder than either alone, and, furthermore, that two sound waves of equal amplitude and period but of opposite phase may combine so as to produce silence. The fact of the interference of sound is conclusive proof that sound is due to wave motion.

In the year 1801 Thomas Young, and fourteen years later Augustin Fresnel, independently demonstrated not only that two lights may combine to produce a light brighter than either one alone, but also that under certain conditions two lights when added together may produce darkness, thereby proving that the phenomenon of interference takes place also in light. Interference, however, is a result which the corpuscular or emissive theory of light absolutely fails to explain, while it is not only consistent with the wave theory, but is demanded by it. It is, therefore, this phenomenon of the interference of light which proves beyond a doubt that light is a vibratory motion of the ether.

If the experiment is tried of allowing the light from two different sources, as *S* and *S′*, Fig. 489, to fall upon a screen at *A*, not the slightest effect of the phenomena due to interference will be observed, but the entire screen will be uniformly illuminated. The reason for the absence of dark regions upon the screen due to interference, even when the sources of light are reduced to mere points, is that a source of light consists of very irregular vibrations, and therefore that the light which is sent out from one source is never in the same phase of vibration as that which comes from the other source. Accordingly, light from both sources reaches every part of the screen in all possible phases, so that destructive interference cannot take place at any one point. On the other hand, if *two beams of light which have originally departed from the same source* are made to meet upon a screen, there will be positions where the light of the two beams reaches the screen in the same phase, and at these places there will be increased illumination, while there will also be regions where the light of the two beams reaches the screen in opposite phase, and at these points there will be darkness. *It is possible, therefore, for interference to take place only when the two interfering beams have originally departed from the same source.*

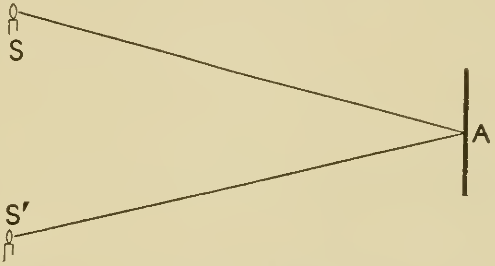

FIG. 489. Light from two separate sources cannot produce interference.

The phenomenon of the production of light and dark bands upon the screen due to interference may be conveniently observed with the aid of the bi-prism, which is a device due to Fresnel. This consists of a piece of glass, polished so as to make a double prism, or *bi-prism, PP′*, Fig. 490, of very obtuse angle. Behind this prism at S is a slit of light, the slit being parallel to the refracting edges. The light emanating from S is refracted and passes through the prism as if it came from the two virtual images, S' and S''. If a screen is placed at some distance from the bi-prism, light will fall upon this screen as if it came from these two virtual sources, and as these are both images of S, they are both in the same phase of vibration. The only purpose of the bi-prism is to provide two sources of light (in this case virtual) very near together, of equal intensity, and in the same phase of vibration. In the further treatment, therefore, we may regard S' and S'', Fig. 491, as two actual sources of light very near together, both in the same phase of vibration, both of the same intensity, and both sending light to the screen. The effect upon every point of the screen will be the algebraic sum of the effects which S' and S'' would produce separately. For the sake of simplicity we will now assume that we are dealing with monochromatic light, or light of one wave length only, such as sodium light.

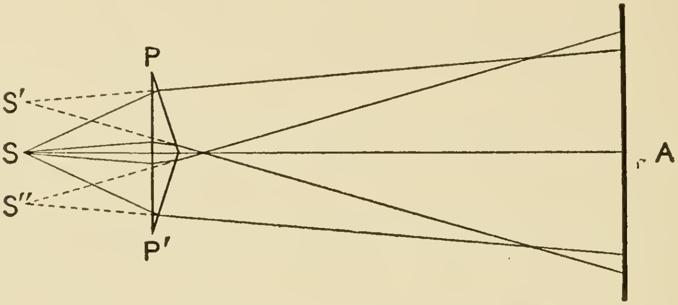

FIG. 490. The Fresnel bi-prism produces two virtual images of the source which are in the same phase of vibration.

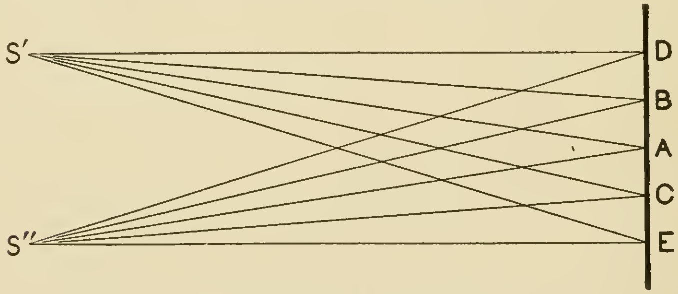

FIG. 491. Two beams of light which have originally departed from the same source are capable of producing interference.

If A is the point on the screen which is equidistant from S' and S'', then

$$S'A = S''A,$$

and the light which reaches A from S' will be in the same phase of vibration as that which reaches this same point from S'', and A will be on the point of increased illumination. If, however, B is a point on the screen such that

$$S''B - S'B = \tfrac{1}{2}\lambda,$$

Where λ is the wave length, the light which reaches B from S'' will have to travel $\frac{1}{2}\lambda$ farther than that which reaches this point from S', and hence the light from these two sources will reach B in opposite phase, and darkness will result. In the same way, if

$$S'C - S''C = \tfrac{1}{2}\lambda,$$

the light which reaches C from S' has had to travel $\frac{1}{2}\lambda$ farther than that which comes to this point from S'', so that light from S' and S'' will meet at C in opposite phase, and C is also a point where destructive interference occurs, and therefore a point of no illumination. Moreover at a point D on the screen, such that

$$S''D - S'D = \lambda,$$

the disturbance which reaches D from S'' is in the same phase of vibration as that which reaches this point from S', and hence D is a point of constructive interference and therefore a point of illumination. For the same reason the point E is a point of illumination, where

$$S'E - S''E = \lambda,$$

and so on. Upon the screen there will therefore appear a series of bands, alternately bright and dark, situated symmetrically upon each side of the point A which is equidistant from the two sources. These light and dark bands are known as interference bands or fringes.

By knowing the distance of the source of light from the screen, and by measuring the distance from the central bright band to the next, that is, the distance AD, the value of $S''D - S'D$, or λ, may be calculated.

In this experiment, if red light instead of sodium light is used, the interference fringes will be formed farther apart, and if blue light is employed, the bands will appear nearer together. This is conclusive proof that red light is of longer wave length than sodium light, while both are of longer wave length than blue light.

Accordingly, if the source is one which emits white light, a central white band will be formed at A, since A is equidistant from S' and S'' for all colored lights, while colored fringes will occur on both sides of this point, the colors appearing in the order of the colors of the spectrum, the blue being always nearest the central white fringe, while the red is farthest from it. A series of six or eight of these colored fringes, each comprising all the colors of the spectrum, may usually be seen on each side of the central white band.

In this, as in all other cases of interference, it is to be observed that no energy, and therefore no light, is lost, but that it is merely redistributed. In other words, the energy from the two sources, which would otherwise produce a uniform illumination upon the screen, is now divided between the light and dark bands, and therefore remains unchanged in amount, although differently disposed. This is shown to good advantage in Fig. 492, which is from a photograph of the interference fringes produced in the manner here described by means of the Fresnel biprism.

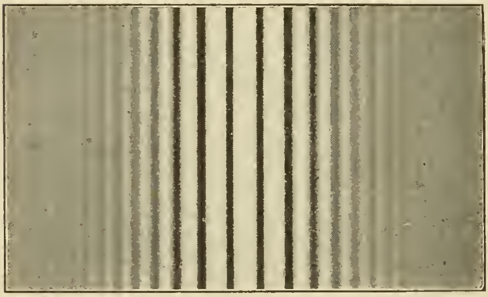

FIG. 492. The interference fringes produced by the Fresnel bi-prism.

669. Interference Due to Thin Films—The Soap Bubble. Whenever a beam of light is incident upon an exceedingly thin transparent film, such as a thin layer of oil on the surface of water, or the sheet of air enclosed between two glass plates, or the coating of oxide which covers the surface of a heated metal, interference of light is produced, and the play of colors seen under these circumstances is due to this phenomenon. These colors may, perhaps, be observed to the best advantage in the thin membrane of water forming the soap bubble, and their production may be explained in the following manner.

Let *MM′NN′*, Fig. 493, represent a highly magnified section of the thin film of water of which the soap bubble is made, and let *AB* be a ray of light incident upon the outer surface. As this ray meets the surface, it is divided into two parts, one of which, *BG*, is reflected, while the other, *BD*, is refracted and enters the film. At the lower surface, this ray *BD* is again divided into two parts, one being refracted and passing out, and the other, *DE*, being reflected at *D*. The ray *DE* is refracted at *E* and passes out of the film in the direction *EF*, a portion, however, undergoing reflection at this point and remaining within the medium. It is with the *refracted* ray *EF* that we have to deal, and this has followed the path *ABCDEF*.

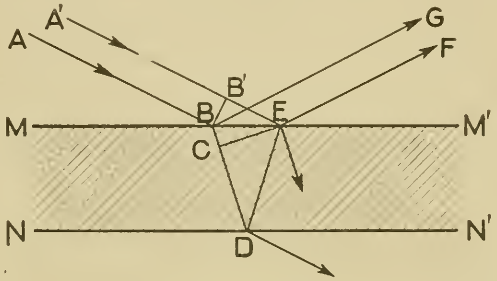

FIG. 493. The interference of light in the film of water forming the soap bubble.

Another ray, *A′E*, parallel to the ray *AB*, will be incident at *E*, and this, like the first, will be divided into two portions, one of which *EF*, will be reflected at *E*, while the other will be refracted and enter the film. In this instance it is the *reflected* portion *EF* of this ray that concerns us further. Since the ray *A′E* is parallel to the ray *AB*, *the ray EF, which leaves the film, is made up of two parts exactly parallel to each other, one of which came originally from the ray AB, while the other is a portion of the ray A′E.* The components of the compound ray *EF*, both of which had their origin in the same source, are therefore capable of interfering; and light or darkness will be produced in the direction *EF* according to the phase in which these two rays unite.

The rays *AB* and *A′E* may be considered as constituent parts of a beam of light of which *BB′* is the wave front in air, and *CE* is the wave front after refraction. That is, while the light has travelled from *B′* to *E* in the air, it has travelled only from *B* to *C* in the denser medium, the ratio of these distances being the index of refraction, *n*. Since the points *C* and *E* are thus in the same phase of vibration, it follows that the component of the ray

EF which came from the ray *AB* has travelled farther than the component which came from the ray *A'E* by the distance from *C* to *D* and from *D* to *E*. But this distance in the film is equivalent to a distance *n* times as great in air, or *n (CD+DE)*, since *n* is the factor of retardation.

At first thought it would seem that if the total optical path from *C* to *D* and from *D* to *E*, that is, a distance in the air equivalent to *n* (*CD+DE*), were exactly half a wave length long, the light which has traversed this path within the film would proceed in the direction *EF* in phase opposite to that of the light which is reflected at *E*. The result of this would then be that the two rays forming the compound ray *EF* would vibrate in opposite phases, and each would destroy the other by interference. It will be observed, however, that the reflection of one of these rays occurs at *D* from a denser to a rarer medium; and therefore, since light waves are transverse, this takes place *without change of sign and without change of phase* (p. 668), while the other reflection occurs at *E* from a rarer to a denser medium, and this then takes place *with change of sign and with a change of phase of half a wave length.*

Accordingly, in estimating the difference of phase which exists between the two components of *EF*, *an additional half wave length must always be added to or subtracted from the difference of phase which already exists due to a difference of distance traversed.* Therefore, if the optical path *CD + DE* in the film, or its equivalent distance *n(CD+DE)* in air, is a half wave length long or any odd number of half wave lengths long, the change of phase of half a wave length which occurs in the reflected ray at *E* will bring the two components of the compound ray *EF* into the same phase, and light will be restored in this direction. But if the optical distance *CD+DE* in the film, or its equivalent distance *n(CD+DE)* in the air is a whole wave length or any number of whole wave lengths long, that is, *if the two components of EF would otherwise be in the same phase due to a difference of path alone, the change of phase of half a wave length which one undergoes by reflection will bring them into opposite phase.* These two components in opposite phase will then mutually destroy each other, and darkness will be produced in the direction *EF*.

Since different colors have different wave lengths, the path

$n(CD+DE)$ will obviously be *an exact number of whole wave lengths for some one color only,* and it is this color alone that will be *destroyed* by interference. On the other hand, that particular color for which the path $n(CD+DE)$ is *an odd number of half wave lengths long will be reenforced* in the compound ray EF. and will appear brilliant in this direction. Therefore, if white light falls upon the soap bubble or any other thin, transparent film in the direction AB, $A'E$, light will be transmitted in the direction EF, and this light will, in general, be colored, the particular tint being a mixture of all the colors that are not destroyed by interference. The colors of the soap bubble and of other thin films are therefore not pure, as are the colors of the spectrum, but consists of all the colors composing white light minus those colors which have thus been destroyed.

If the film $MM'NN'$ forming the bubble were infinitely thin, so that the ray which traverses it along the path CDE had practically no farther to travel than the one which is simply reflected from the surface, that is, if the film were so thin that the path CDE is practically zero for all wave lengths, then the difference of phase due to the inequality of path alone would be negligible. The two components of EF, however, would still be in opposite phase, due to the fact that one had undergone reflection from the exterior and the other from the interior surface of a denser medium, and therefore one component would have suffered a change of phase of half a wave length while the other would not. Destructive interference would therefore take place for all colors, and *the film when seen by reflected light, would appear black.*

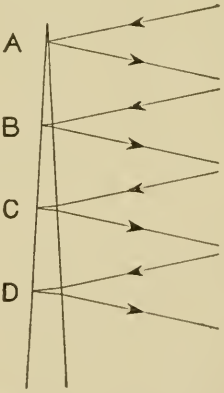

FIG. 494. Interference of light in a wedge-shaped film.

If a soap bubble film is formed on a wire ring two or three inches in diameter and held vertically, gravity will cause it to settle in a wedge-shaped film, Fig. 494, the thin edge of the wedge being uppermost. If white light falls upon this film of varying thickness, part will be reflected from the first surface, while part will traverse the film and be reflected from the rear. If these two beams are in the same phase, light will be produced, but if they are in opposite phase light will be destroyed. Destructive interference occurs when the optical path of the ray within the film is zero, or a whole wave length, or two whole wave lengths, and so on, for then the two components are in opposite phase, due to the change of phase of half a wave length which has occurred by reflection from the rarer to the denser medium. A play of colors will therefore, be seen in this wedge-shaped film, the colors being arranged in horizontal bands, and continually changing in tint as the film grows thinner.

Ultimately, if the film is a good one, the thin edge will have a thickness which is vanishingly small as compared with the length of a wave of light, and the upper portion of the wedge will then appear black, since at this point it is of zero thickness for all wave lengths. At points farther down, as at *B, C, D,* etc., the film will be exactly one wave length thick, two wave lengths thick, three wave lengths thick, etc., for some one color, hence at these points that color will be destroyed by interference, and the light which will be seen is a mixture of all the other colors that make up white light except the one thus destroyed. There will thus be seen a series of colored spectra arranged in parallel bands, the colors, however, changing in a most beautiful manner as gravity and currents of air cause the thickness of the film to vary.

670. Newton's Rings. In Newton's rings we have another beautiful illustration of the production of color due to the destructive interference of light. A convex lens of long focal length, when placed upon a plane surface of glass, Fig. 495, encloses, between the glass plate and itself a thin film of air of varying but determinate thickness. When white light falls upon this system, part of the light which enters the lens is reflected from the lower curved surface, and part is reflected from the upper surface of the plane glass plate. Thus, a ray of light *AB* will be divided into a reflected ray *BG,* and into the refracted ray *BD.* The latter is reflected at *D* and *again enters the lens at E, passing out along EF.* At *E* another ray, *A'E,* parallel to the first is incident upon the lower surface of the lens and is divided into two parts, one of which is reflected, while the other is refracted. The *reflected* portion of this second ray, which passes likewise along *EF*, *coincides in direction with that portion of the first ray which has traversed the thin layer of* air *BDE*. Interference, therefore, takes place between the two component parts of the ray *EF*, and that color for which the difference of phase of these two components is half a wave length is destroyed by interference.

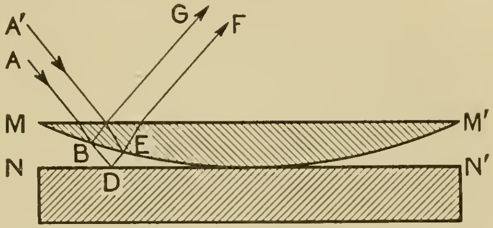

FIG. 495. Interference of light in a film of air, producing Newton's rings.

By drawing upon a larger scale that portion of the diagram through which these particular rays pass, Fig. 496, we see, as before, that, while *BB'* is the wave front of the incident beam in the lens, *CE* is the position which this wave front assumes in the air-film between the lens and the plate. In other words, light travels from *B* to *C* in the air in the same time in which it travels from *B'* to *E* in the lens. The points *C* and *E* are therefore in the same phase of vibration, and, accordingly, one of the two components of the compound ray *EF* has travelled the optical distance from *C* to *D* and from *D* to *E* farther than the other. If no change of phase due to reflection had taken place in either of these components, and therefore if the only difference in their phase had been due to a difference of path alone, it is evident that if the distance *CD* +*DE* were a whole wave length or any number of whole wave lengths long, light would be reenforced in the direction *EF*; while if this distance were a half wave length, or any odd number of half wave lengths long, light would be destroyed in this direction by interference.

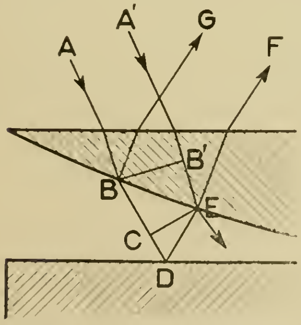

FIG. 496. Details of the production of color in Newton's rings.

Since one reflection, however, is from a denser to a rarer medium, while the other is from a rarer to a denser medium, one of these components has changed its phase with respect to the other by half a wave length, this change undoubtedly taking place in the reflection from the rarer to the denser medium at *D*. Therefore, *if the path CD+DE in the air-film is a half wave length or any odd number of half wave lengths long*, the change of phase of one of these rays by half a wave length will bring the two components of *EF* into the *same phase*, and light will be *produced* in this direction. But *if the path CD+DE is zero, or a whole wave length, or any number of whole wave lengths long*, then the change of phase of half a wave length which one ray has undergone by reflection at *D* will bring the two components of *EF* into *opposite* phase, and light in this direction will be *destroyed*. Obviously this difference of path will be a whole wave length for some one color only; hence, if white light is used, this color alone will be destroyed by interference, and the resulting beam *EF* will be colored, the tint being a mixture of all the colors which make up white light except the one which has thus been destroyed

At a greater distance from the center of the lens the thickness of the air film increases, hence light of longer and longer wave length is destroyed in this way. Accordingly, the light which is propagated in the direction *EF* changes in color as we recede from the center of the lens, until a point is found where the length of the path *CDE* is two wave lengths, and at this point a mixture of undestroyed colors will be produced which is the same as that previously obtained when this difference of path was only one wave length. Inasmuch as the thickness of the air film increases nearly uniformly between these two points, each color throughout the spectrum will be successively destroyed by interference, and the colors transmitted to the eye from intermediate points will assume various tints which, although they do not consist of one wave length only, suggest the colors of the spectrum. At a distance farther from the center, a point will be found where this difference of path is three wave lengths, when the same phenomenon will be repeated, and so on. Colored bands resembling spectra are therefore formed at gradually diminishing distances from each other; and inasmuch as a certain thickness of the air film is uniform in a circle about the center of the lens, these bands of colored light are circular in form, Fig. 497.

When seen by reflected light, the center of this system of rings is black, since here the lens, which indeed presses upon the glass plate, is separated from the latter by a film of air, the thickness of which is exceedingly small as compared even with the length of a wave of light. The difference of path being then practically zero, the only retardation produced between the two components of the ray *EF* is the change of phase of half a wave length which one component suffers by reflection. Inasmuch, however, as this change occurs here for all wave lengths, it follows that destructive interference for all colors takes place, and *the center of the ring system is black.*

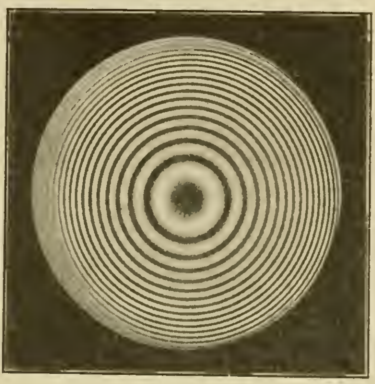

FIG. 497. Newton's rings in monochromatic light.

671. The Use of the Wave Length of Red Cadmium Light as the Standard of Length. By means of an exceedingly ingenious device known as the interferometer, Michelson has been able to apply the phenomenon of the interference of light to the measurement of linear distances. The main features of this instrument and the principle upon which its operation depends may be briefly described as follows:

The light from a source *S*, Fig. 498, falls upon a plate of glass *ab*, which is coated with a film of silver so thin that it transmits practically as much light as it reflects. The incident beam *SI*, falling upon this half-silvered mirror at *I*, is then divided into two parts, one of which, the reflected beam, passes back through the glass and is refracted to the mirror *M*, and the other, the transmitted beam, passes on from *I* to the mirror *N*. The mirrors *M* and *N* are placed normally to these beams respectively, so that the light which falls upon them is reflected back upon itself. That is, the light which falls upon *M* is reflected back to *I* and is then transmitted to the eye at *E*, while that which falls upon *N* is likewise reflected back to *I* and again reflected to the eye at *E*. In the path of the latter beam is placed the plate of glass *cd*, which is unsilvered but otherwise exactly like *ab*, so that each of these beams may pass through equal thicknesses of glass. Accordingly, the beam of light between *I* and *E* consists of two parts, one of which has traversed the distance from *I* to *M* and back again to *I*, while the other has traversed the path from *I* to *N* and back again to *I*. Inasmuch as the two beams of light between *I* and *E* originally departed from the same source, they are capable of interfering, and will produce light or darkness according to the phase in which they meet.

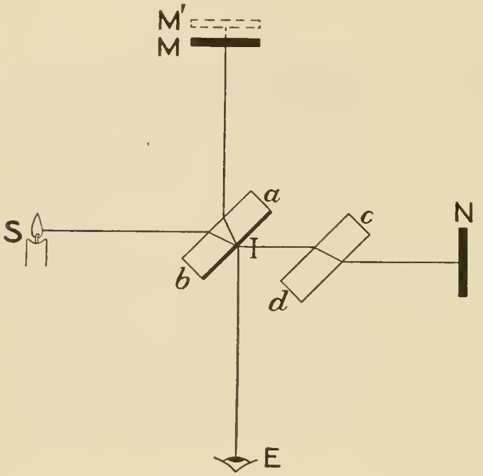

FIG. 498. Diagram showing the principle of the interferometer.

If the distance from *I* to *M* and back again to *I* is exactly equal to that from *I* to *N* and back to *I*, destructive interference will occur in the beam *IE*, and darkness will be produced at *E*. At first it would seem that these two beams would be in the same phase, and therefore would produce reenforcement, since the lengths of the two paths are identical. We see, however, that the beam which passes from *I* to *M* is reflected at *I* from a denser to a rarer medium, and so undergoes no change of phase, while that which returns from *N* is reflected at *I* from a rarer to a denser medium, and so suffers a change of phase equal to half a wave length. This brings the two beams between *I* and *E* into opposite phase, and destructive interference occurs.

If, now the distance from *I* to one of these mirrors, *M*, for instance, is increased by one-quarter of a wave length, so that the beam which traverses this distance in going and returning will lengthen its path by half a wave length, the two interfering beams between *I* and *E* will then be in the same phase, so that constructive interference will take place and light will be restored. The further increase of this distance by the same amount will bring the two beams again into opposite phase, and darkness will result, while an additional increase of this distance, again by the same amount, will once more produce illumination, and so on.

This succession of light and dark bands may therefore be made to pass before the eye by so mounting the mirror at *M* that, while its plane may be kept always parallel to itself, its distance from *I* may be gradually increased. Minimum light will prevail at *E* when the path from *I* to *M* and back again to *I* is exactly equal to that from *I* to *N* and back to *I*, or when the former path exceeds the latter by a whole number of wave lengths; while illumination will be produced at *E* when the light, in passing from *I* to *M* and back again to *I*, travels half a wave length or any odd number of half wave lengths farther than that which traverses the distance from *I* to *N* and back to *I*. Accordingly, by counting the number of interference fringes which occur at *E*, and by carefully measuring the distance through which the mirror *M* has been moved, the number of wave lengths of the particular light employed which are equivalent to this linear distance may be accurately determined.

After an examination of the light from many sources, Michelson found that, of the three bright emission lines in the spectrum of the vapor of cadmium, the one in the red is composed of light which is monochromatic, that is, it consists practically of one wave length only. It was then apparent that, by producing interference between two beams of this light in the method described, and by counting the number of interference fringes which passed before the eye as the mirror *M* was moved through a measured distance, it would be entirely possible to express this distance as equivalent to a certain number of wave lengths of this particular kind of light. This Michelson has done with the meter, and has found that the international meter prototype (p. 13) is equivalent in length to 1,553,163.5 wave lengths of red cadmium light.

A matter of far greater importance, however, is attached to this investigation than the mere determination of the number of wave lengths of a particular light which may be equivalent to a given distance. Inasmuch as the wave length of red cadmium light depends upon the period of the electronic vibration within the cadmium atom, it follows that this wave length is a distance which is absolutely fixed and definite. Moreover, this distance, though small, is indestructible and unchanging with time, and these are the essentials which are demanded for a standard of length (p. 12). These conditions are met by no other known distance, so that practically this makes the wave length of red cadmium light our future standard of length. Instead, therefore, of depending for a standard upon the length of a particular bar, with all the attendant possibilities of injury or destruction, we have here, in the wave length of red cadmium light, a standard of length which fulfills every requirement, in that it is absolutely invariable, it persists for all time, and is always and everywhere accessible.

672. The Rectilinear Propagation of Light. Let *S*, Fig. 499, be a source of light, and *mn* a portion of a spherical wave departing from *S* as a center. Let *P* be a point which is to be illuminated by the light which proceeds from *S*. We are to inquire as to what part of the wave front the illumination at *P* is due.

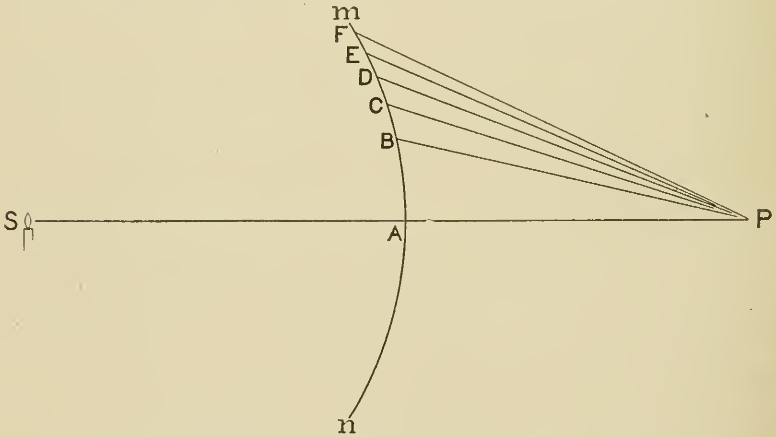

FIG. 499. Each element of the wave front sends light to a distant point.

In our discussion of the Huygens construction for the propagation of a spherical wave (p. 733), we have seen that every particle upon the wave front must be considered as a center of disturbance sending out spherical waves precisely similar to those which depart from the source of illumination at the center. Accordingly, every point upon the wave front *mn* is in vibration and is sending out little spherical waves; and the result at *P*, that is, the total illumination produced at this point, is the combined effect of the waves which are thus sent to *P* from all points upon the wave front.

If *PA* is the perpendicular let fall from *P* upon the wave front, we can divide the latter surface into sections *AB, BC, CD, DE,* etc., such that the distance from *P* to *B* is half a wave length longer than that from *P* to *A*, the distance from *P* to *C* is half a wave length longer than that from *P* to *B*, and so on. Since all points on the wave front are in the same phase of vibration, it is evident that the light which reaches *P* from *B* is half a wave length behind that which reaches this point from *A*, since it has had half a wave length farther to travel. Therefore, the light which comes to *P* from *B* will arrive opposite in phase to the light which reaches *P* from *A*, and hence the illumination at *P* which the light from *A* would produce is destroyed through interference by the effect at this point which the light from *B* would produce. Similarly, the light from *C* will reach *P* opposite in phase to the light from *D*, the light from *D* will arrive at *P* opposite in phase to the light from *E*, and so on. Moreover, every point in the section *DE* is sending out light which will reach *P* opposite in phase to the light which is sent out from a corresponding point in the section *CD*, and therefore if *CD* and *DE* were equally large, the entire disturbance which is sent out from *CD* would be destroyed at *P* through interference by the disturbance which is sent out from *DE*.

In the same way, for every point in the section *BC* there is a corresponding point in the section *AB* which will send out light that will reach *P* in opposite phase. Hence, if the sections *AB* and *BC* were equally large, the effect at *P* of the light from the section *BC* would be entirely destroyed through interference by the light which reaches *P* in opposite phase from the section *AB*. It is evident, however, that these sections are not of equal size, but that the section *AB* is larger than the section *BC*, *BC* is somewhat larger than *CD*, *CD* is slightly larger than *DE*, while at a very short distance from *A* the sections rapidly become practically equal. As a matter of fact, the section *CD* is less than 1/10 larger than the section *DE*, so that less than 1/10 of the section *CD* is of service in sending light to the point *P*. Also, the section *BC* is only about half as large as the section *AB*, and therefore while half of the section *AB* is rendered ineffective in sending light to *P*, due to interference from the section *BC*, the remaining half of this section is thus operative, and *P* is therefore illuminated.

Accordingly, the illumination at *P* is due almost wholly to about half of the first half wave section or half period element immediately about the point *A*, while the light which would otherwise reach *P* from the remaining portions of the wave front is entirely destroyed by interference. If the point *P* is three meters distant from *A*, the region of the wave-front which is effective in sending light to *P* is probably not more than 1 millimeter in extent. This small area about the point *A* to which the illumination at *P* is almost entirely due, is called *the pole of the wave.* It is to be observed that the pole of the wave is on the perpendicular let fall from the point *P* upon the wave, and therefore the line *SAP* is a straight line. That is, it is only the portion of the light emanating from *S* and proceeding in the direction *SA* that is effective in producing illumination at the point *P*, or, in other words, *the light which reaches P is propagated in a straight line from the source S.* We see, therefore, that the light which illuminates *P* cannot follow the paths *SBP, SCP, SDP,* and so on,

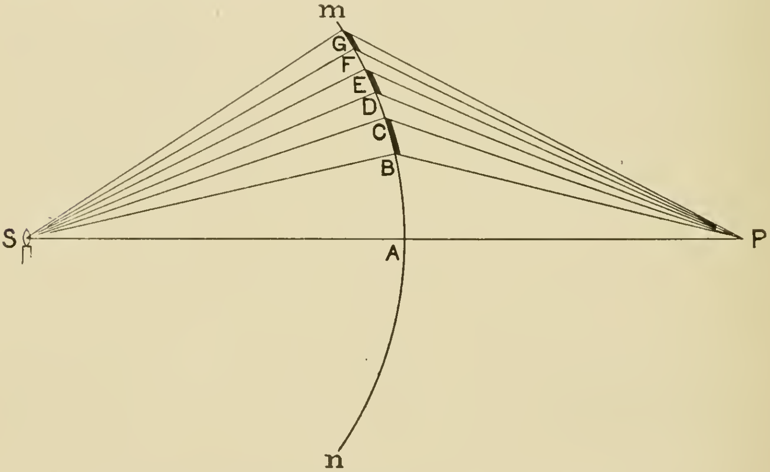

FIG. 500. A series of small screens, by intercepting portions of the wave front, prevent the interference which would otherwise take place.

but that it advances in the straight line SAP. From this it is evident that *light travels in straight lines because of interference.*

If small screens, Fig. 500, are placed at BC, DE, etc., the interference which results from these portions will be prevented, and the light at P will be greatly increased. In this case the light no longer passes only in straight lines from S to P, but also proceeds from S to P by the paths SCP, SDP, and so on. Therefore, *when interference is prevented, light does not travel in straight lines.*

673. The Principle of Least Time. The important truth which is here presented, namely, that *light travels in straight lines because of interference,* shows, furthermore, that the light which is effective in illuminating P is that which reaches this point *in the shortest time* from the wave front. This is but one instance of a very general law, the statement of which is due to Fermat, that *when light passes from one point to another,* even though it undergoes many reflections and refractions, *it pursues the path by which it can travel in the least time.* Moreover, the fact that the light which illuminates a point proceeds to that point only from the pole of the wave is conclusive proof that in a homogeneous isotropic medium, that is, one possessing the same properties in all directions, light is propagated in a direction normal to the wave front.

674. Diffraction. If a screen XY, Fig. 501, is now so placed as to intercept that portion of the wave front below A, the point P will be almost entirely unaffected, provided only that this screen does not intercept the half period element about A. The line AP will therefore mark the boundary of the geometrical shadow, the region above this line being illuminated and the part below being in the shadow. A careful examination of the boundary of the shaded portion will show, however, that the shadow is not perfectly sharp, even though the source of light S be a single point, but that *some of the light does bend around the edge of the screen XY and proceeds into the shadow,* so that a point P′ a little below the line AP is faintly illuminated.

If we consider the point P', Fig. 502, within the geometrical shadow, but very near the line AP, we shall see how it is that some of the light from the wave front does bend into the shaded region and is effective in faintly illuminating *P′*. If we draw a perpendicular *P′A′* from *P′* to the wave front, and then, as before, lay off points, *A′*, *B′*, *C′* *D′*, *E′*, upon the wave front which are successively half a wave length farther from *P′*, *P′* will be brightly illuminated if the screen *XY* does not intercept the first half period element *A′B′*. But if the screen intercepts this first half period element, as in Fig. 502, then *P′* would be in total darkness and the shadow would be perfectly sharp in the half period element *B′C′* were exactly equal to the half period element *C′D′*, and if *D′E′* were exactly equal to *E′F′*. In that case,

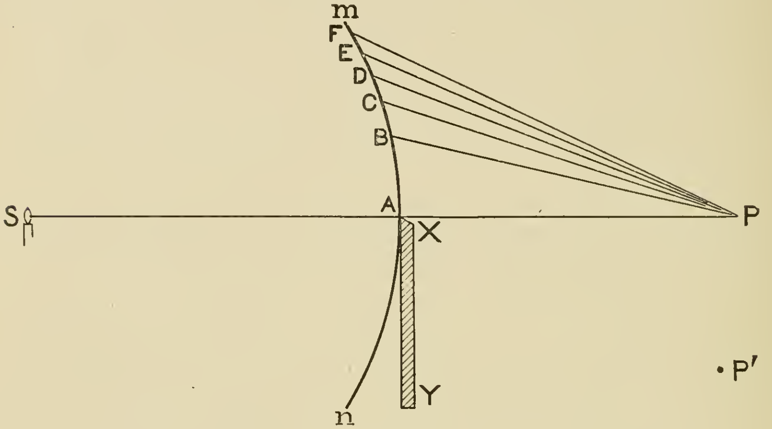

FIG. 501. The illumination of a point on the boundary of the geometrical shadow.

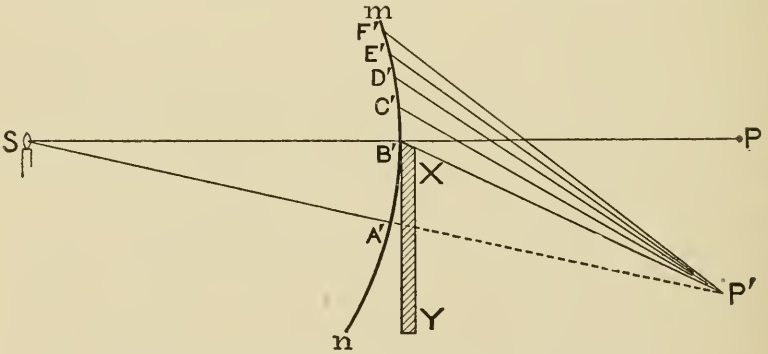

FIG. 502. The edge of the geometrical shadow is not perfectly sharp, due to a slight bending of the light around the edge of an object.

the light which would reach P' from $B'C'$ would be entirely destroyed through interference by the light which would reach this point from $C'D'$ and the light from $D'E'$ would destroy at P' the light from $E'F'$. But $B'C'$ is about one-fifth greater than $C'D'$, hence there will remain about one-fifth of $B'C'$ which is not destroyed by the interference of $C'D'$, and hence this small part of $B'C'$ will be effective in sending light to P'.

In the same way, inasmuch as $D'E'$ is slightly larger than $E'F'$, there will be a very small part of $D'E'$ which is not destroyed through interference by $E'F'$, and therefore this small undestroyed part of $D'E'$ will be effective in sending light to P'. Therefore, if P' is very close indeed to the boundary of the geometrical shadow, it will be faintly illuminated, not, indeed, by light from the pole of the wave, but by the light which proceeds from the undestroyed part of the three or four half-period elements immediately above the pole. As P' proceeds farther into the geometrical shadow, and the obliquity of the lines drawn from the edge of XY to P' increases, the intensity of the light at P' decreases very rapidly, and therefore, at a very short distance from the boundary of the geometrical shadow, P' is in a region of total darkness. It is thus obvious that *the outline of a shadow is not perfectly sharp, but that light really does, to a very slight extent, bend around the edge of an obstacle and penetrate into the geometrical shadow.*

This deviation of light from a straight line as it passes the edge of an opaque body, and its propagation, although to a limited degree, into the region of the shadow, is called diffraction. Because of the phenomenon of diffraction, therefore, the boundary of a shadow is not perfectly sharp, and hence it is true that *light does not travel strictly in straight lines.* Likewise because of this phenomenon, an illumination does exist for an exceedingly short distance within the geometrical shadow, so that the intensity of the light diminishes very rapidly, though not abruptly, from the boundary of the geometrical shadow to the region of total darkness.

675. **Young's Interference Experiment.** The celebrated experiment of Thomas Young, which proved to be the crucial test of the wave theory of light, depended in reality upon the phe-

nomenon of diffraction, or the bending of light into the geometrical shadow.

Through a minute opening, *S*, Fig 503, Young admitted sunlight into a darkened room, and in the slender beam of light so formed he placed a screen with two pinhole-openings, *S′* and *S″*. These two openings were thus upon the surface of a spherical wave front of light which had departed from S as a center, and so were in the same phase of vibration. Accordingly, spherical wavelets, also in the same phase of vibration, then emanated from these openings as centers, and advanced to the distant screen *MN*. Here two bright spots, *I′* and *I″*, were produced, these being, indeed, the direct illuminations from *S*, since *S′* and *S″* were upon the poles of the waves which proceeded to *I′* and *I″* respectively. Furthermore, by the above phenomenon of diffraction, the light from *S′* and *S″* underwent a bending at the edges of these openings, and penetrated into the shadow cast by the opaque screen.

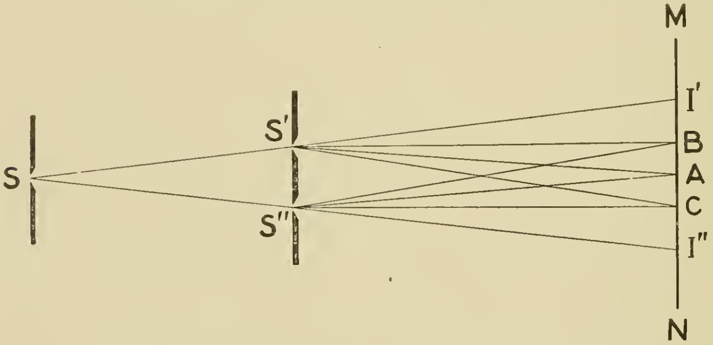

Fig. 503. Young's interference experiment, demonstrating the wave theory of light.

Between the points *I′* and *I″*, Young observed a series of light and dark bands, which he interpreted correctly as being the points at which the waves from *S′* and *S″* met in the same or in opposite phase. The point *A*, midway between *I′* and *I″*, was a point of illumination, since it was equidistant from *S′* and *S″*, and the waves which departed from these openings met at *A* in the same phase. Similarly, *B* and *C* were points of darkness, since the light from one of the openings had to travel half a wave length farther to each of these points than light from the other opening. Therefore, the light from the two openings met at these points in opposite phase, and each destroyed the other by mutual interference. Other light bands were likewise found at points on the screen which differed in distance from the two openings by one wave length, two wave lengths, and so on, that is, by an *even* number of half wave lengths, while other dark bands were also found at points which differed in distance from the two openings by an *odd* number of half wave lengths.

By intercepting the light from either one of the openings, the bands instantly disappeared, thus proving conclusively that the alternations of light and darkness which had been observed had been points of reenforcement or destruction, due to the effect of the light from the opening which was now obscured. The phenomenon of destructive interference, however, was one which the corpuscular theory of light failed completely to explain, while it was, as Young proved, fully interpreted, and, indeed, demanded, by the wave theory. Notwithstanding this demonstration, the wave theory of light for an entire generation still failed of general acceptance.

676. The Diffraction Grating. The series of small screens (p. 860) which serve to intercept portions of the wave front and thus to prevent interference, is called a *diffraction grating,* or, more simply, *a grating.* Inasmuch as these screens intercept a portion of the wave front which would ordinarily produce interference, the rectilinear propagation of light can no longer be maintained, and *light is transmitted into regions which would otherwise be within the geometrical shadow.*

Since, in general with the grating, it is important to use a plane wave front, the screens are placed upon a plane surface and are equidistant. Formerly they consisted of a number of carefully arranged wires, separated from one another by equal spaces; but in later years they are made by ruling upon a piece of glass or polished speculum metal a series of *fine, parallel, equidistant lines,* at the rate of several thousand to the centimeter. The metallic grating then constitutes what is known as the *reflection grating,* while the one ruled on glass, which is the only one we shall here consider, is called the *transmission grating.* Although it is undoubtedly true that the scratches upon the glass are not perfectly black, they serve at least to interrupt the continuity of the wave front, and to destroy the uniformity of phase which there exists. These scratches may therefore be regarded as a series of opaque, equidistant screens, which intercept portions of the wave front, and which, accordingly, prevent interference in certain directions.

The prevention by the grating of destructive interference, and its action in transmitting light into regions which would otherwise be in darkness, may be explained with the aid of the following diagrams. Fig 504 shows a highly magnified cross-section of the ruled surface, the grooves or furrows of which must be conceived as drawn perpendicular to the plane of the paper. The *width* of each groove is represented by the heavy lines *AB, CD, EF,* etc., while the width of the clear spaces between the scratches is indicated by the openings *BC, DE, FG,* etc. The grooves, or furrows, thus serve as the opaque screens which are to intercept portions of the wave front, thereby producing the phenomenon of diffraction in the manner to be described, while the alternate spaces are the openings through which light passes from the distant source at the left. *mn* represents a wave front advancing towards the grating from a source so distant that the rays, which are indicated by arrows, may be considered as parallel, and the wave front may be regarded as plane. Indeed, the vertical lines to the left may be considered as a succession of wave fronts, each of which is distant from the next by one wave length of light. Inasmuch as all points upon the wave front consist of particles which are executing similar vibrations at the same time, it follows that *the surface of the grating is continually the locus of particles which are in the same phase of vibration.*

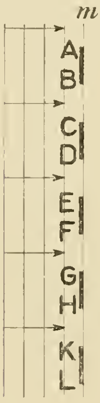

Fig. 504. The diffraction grating.

Therefore, by Huygens's construction for the propagation of a wave front (p. 733), we see that every point of each opening must be considered as a center of disturbance sending out little spherical wavelets which are all in the same phase of vibration, and that the envelope of all these little wavelets is the new wave front. Inasmuch as it is impossible for us to consider *every* point in each opening, we shall here select only what may be called *corresponding points*, that is, points which are similarly situated in each aperture, such as the point A in the first opening, C in the second, E in the third, and so on. The distance between corresponding points, which, indeed, is the distance between the rulings on the grating, is known as the *grating space*.

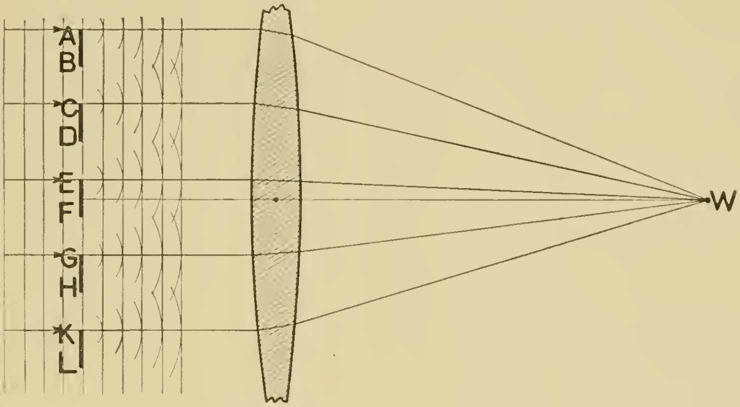

FIG. 505. The formation of the white central image.

Accordingly, when the vibration which originates at A has travelled one wave length, Fig. 505, this disturbance is found upon the surface of a little spherical wavelet which has one wave length as a radius. Inasmuch, however, as the vibration at C is in the same phase as that at A, the disturbance which goes out from C is likewise found upon the surface of a little spherical wavelet, the radius of which is one wave length. The same holds true also for E, for G, and for the corresponding point of every opening of the grating. The envelope of the little spherical wavelets gives a new wave front on the other side of the grating, so that at least a portion of the light which passes through the openings continues as a plane wave, the wave front being parallel to the grating surface and to the wave fronts before reaching the grating. Moreover, if the above is true for light of one color, and therefore for one wave length, the same is likewise true for light of another color and a different wave length. If, then, *white* light advances towards the grating, the wave fronts of lights of different colors will still remain parallel to themselves after passing through the grating, and the light which is thus transmitted by the grating still remains white.

If, now, a converging lens is placed with its principal axis perpendicular to the grating and to the advancing wave fronts, it will converge these wave fronts to a point at its principal focus, where they will all meet in the same phase, the greater thickness of the central portion of the lens just compensating for the shorter distance. In other words, the optical distance from every point on a plane wave front through the lens to the focus is the same. The first effect of the grating, therefore, is to produce a white spot of light, known as the *white central image*, on the principal axis of the lens, and at its principal focal distance.

It is possible, however, to form another envelope of the wavelets which emanate from corresponding points of the grating openings, as may be seen in Fig 506. Since the points *A*, *C*, *E*, *G*, etc., are in the same phase of vibration, it follows that, after the disturbance which leaves *A* has advanced exactly one wave length, it is then in the same phase of vibration as the *next* disturbance which is *about* to leave *C*. Accordingly, the line drawn from *C* tangent to the wavelet at *c* represents the envelope of points which are in the same phase of vibration. In a similar manner, after the disturbance from *A* has advanced exactly two wave lengths, that from *C* has travelled one wave length, and that from *E* is about to start. Thus the line *Ec*, which is the envelope

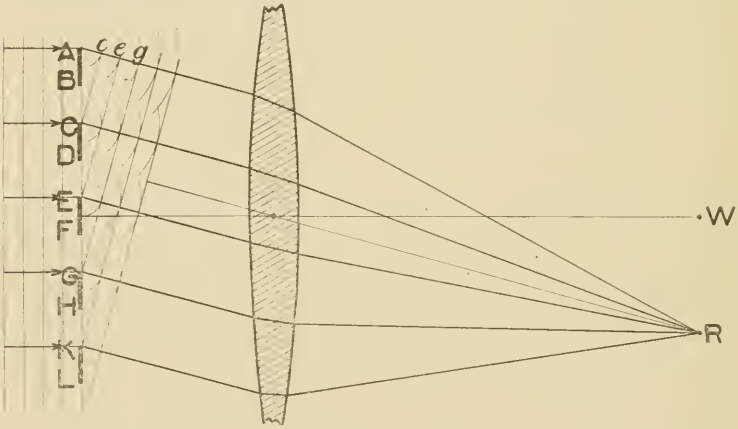

FIG. 506. The production of red light in the spectrum of the first order.

of these points which are in the same phase of vibration, is likewise a wave front. So, also, is *Gg* a wave front, and so on, and we see that light is propagated perpendicularly to these wave fronts.

The light, however, which is transmitted in this direction is not white, since the distance *Ac*, which is exactly one wave length, *can be a wave length for some one color only.* Therefore, since lights of different colors have different wave lengths, *light of one color only,* such as red, is transmitted in a direction perpendicular to the system of wave fronts *Cc, Ee, Gg,* etc. The converging lens, still remaining with its principal axis perpendicular to the plane of the grating, will thus bring light of this particular color to a focus at *R* upon a secondary axis, that is, upon a line through the optical center of the lens perpendicular to the advancing wave fronts.

Furthermore, if we construct a similar diagram for violet light, which, as we know, is of shorter wave length than red, Fig. 507, we shall find that, as *Ac* is now shorter, the wave fronts *Cc,*

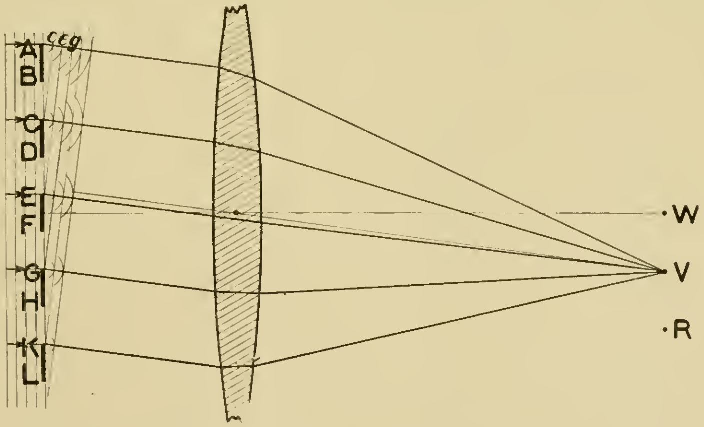

FIG. 507. The production of violet light in the spectrum of the first order.

Ee, Gg, etc. are inclined at a smaller angle to the surface of the grating. The lens will then bring these wave fronts to a focus at the point *V*, which is on the secondary axis perpendicular to the wave fronts, where *V* lies nearer than *R* to the white central image. Those colors whose wave lengths are intermediate between violet and red will then be brought to foci at points intermediate between V and R; so that if white light falls upon the grating, it is resolved into its component colors, and a spectrum is produced. It is to be observed, however, that *the colors in the grating spectrum are arranged in an order which is the reverse of those in the prismatic spectrum; for while violet in the prismatic spectrum is deviated most, this color in the grating spectrum is deviated least.* Moreover, the deviation of a particular color is seen to depend upon the distance Ac, that is, upon the wave length of this particular colored light. Accordingly, the various colors of the spectrum VR, which are thus seen to be found in the principal focal plane of the converging lens, are distant from the white central image by amounts which are exactly proportional to their various wave lengths. On this account, the grating spectrum, or the diffraction spectrum, as it is called, is also known as a *normal* spectrum.

The spectrum here described is known as *the spectrum of the first order.* It is possible, however, also to construct another envelope of wavelets which are in the same phase of vibration, as these pass through corresponding points of the grating. By referring to Fig. 508, we see that after the disturbance which leaves

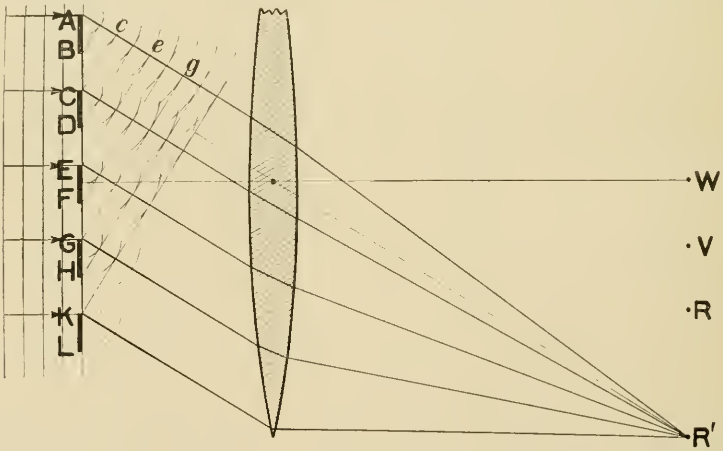

FIG. 508. The production of red light in the spectrum of the second order.

A has advanced *two* whole wave lengths for light of some one color, such as red, it will then likewise be in the same phase of vibration as the disturbance which is about to leave C, so that the envelope of these wavelets forms a new wave front for red light. Also, after the disturbance which leaves A has advanced four wave lengths, that from C has travelled two wave lengths, and that from E is about to start. These three wavelets are then in the same phase of vibration, and their envelope constitutes a new wave front for red light, and so on.

Accordingly, we see that at an angle greater than that which was shown in Fig. 506, red light will again be restored by the grating; and also, at an angle greater than that shown in Fig. 507, violet light will likewise be restored. Colors intermediate between these two will then be transmitted to intermediate points, the whole forming what is known as *the spectrum of the second order.* The angular deviation of this spectrum from the white central image is thus seen to be greater than that of the spectrum of the first order. In a similar way, spectra of the third, the fourth, and even the fifth orders may be formed. Moreover, since the wave fronts might have been drawn with inclinations upwards as well as downward, it is seen that a similar series of spectra, also in the focal plane of the lens, but on the opposite side of the white central image, will be formed.

It is obvious that, without the grating, the only effect of the advancing wave front, as it passes through the lens, would be to produce this white central image. In this case, each portion of the wave front would destroy by interference the light which would otherwise be transmitted from adjacent portions of the wave front to points which are now in the shadow. When, however, the grating is in position, its ruled lines intercept those portions of the wave front which would produce interference, so that light ceases to travel in straight lines, and is propagated into regions which would otherwise be without illumination.

In addition to the advantages which the grating possesses in producing a normal spectrum of high dispersion, is the readiness with which this device lends itself to the measurement of the wave length of light. Let us, for the present, confine our attention to the spectrum of the first order, and consider the wave fronts of light of a particular color, such as red, Fig. 509. If these wave fronts are inclined at the angle θ_1 to the surface of the grating, we then know that light of this color is propagated in a direction at right angles to these wave fronts, and therefore at the angle

θ_1 with the direction in which the white light which forms the white central image is propagated. This, however, is possible only when the perpendicular from *C*, let fall upon the ray which leaves *A*, intercepts the distance *Ac*, *which is exactly one wave length, or* λ. *for light of this particular color.* Furthermore, if *l* is the distance between corresponding points of the grating, or the *grating space,* we see that

$$\frac{\lambda}{l} = \sin \theta_1,$$

or

$$\lambda = l \sin \theta_1.$$

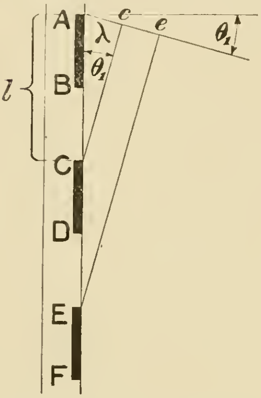

Fig. 509. The use of the grating in the measurement of the wave length of light.

For the spectrum of the second order, Fig. 508,

$$\frac{2\lambda}{l} = \sin \theta_2,$$

or

$$\lambda = \tfrac{1}{2} l \sin \theta_2$$

and for the spectrum of the n the order

$$\frac{n\lambda}{l} = \sin \theta_n,$$

or

$$\lambda = \frac{1}{n} l \sin \theta_n.$$

Here θ is the angle by which the light of any particular wave length is deviated from its original direction. *l* may be determined by measuring with a microscope the distance from groove to groove upon the grating, or, more accurately, by counting the number of lines to the centimeter when the grating is ruled. This number is always given with the grating, and is known as *the grating constant*. Since the two quantities upon which λ depends are now known, the wave length of this particular colored light may be determined with great accuracy.

Measured in this way, the wave length of sodium light is found to be approximately

$$\lambda = 0.0005896 \text{ millimeters.}$$

With the knowledge of the wave length of light, and with the aid of the fundamental equation for the propagation of wave motion (p. 679), namely,

$$n\lambda = V,$$

where V is the velocity of light, it is possible also to determine n, of the number of vibrations per second. Remembering that

$$V = 300{,}000 \text{ kilometers per second,}$$

or

$$V = 300{,}000{,}000{,}000 \text{ millimeters per second,}$$

we see that

$$n = \frac{300{,}000{,}000{,}000}{0.0005896},$$

from which

$$n = 508{,}800{,}000{,}000{,}000,$$

or approximately 500 million million, which is the number of vibrations per second made by sodium light.

CHAPTER LVII.

THE POLARIZATION OF LIGHT

677. Polarization by Reflection. In the preceding treatment of light, including reflection, refraction, dispersion, and interference, no reference has been made as to the direction of vibration of the particles which constitute the ether waves. Indeed these various phenomena might be brought about equally well by longitudinal or by transverse vibrations, or by vibrations which are confined to a single plane or by those taking place in many planes. It is, therefore, by experiments of a nature different from the more familiar ones heretofore considered that all questions which have to do with the character of the ether vibrations, as well as the nature of the waves to which these vibrations, give rise, must be answered.

From the fact that a beam of ordinary light may be reflected or refracted or may suffer interference in one direction as well as in any other, it is evident that this beam of light is alike on all sides, or that it has properties on one side which are precisely the same as those on any other. These effects might be produced equally well, so far as these alone are concerned, if the vibrations constituting this beam were longitudinal or if they were transverse, provided, however, in the latter case, that the transverse vibrations were to take place in all possible planes, so as to present conditions which are absolutely the same on every side of the ray. It is, therefore only when a beam of light can be made unsymmetrical, that is, when it can be so altered as to have properties on one side which it does not possess on another, that conclusions may be drawn as to the character of the vibrations of which this beam is composed. Such a lack of symmetry may indeed be produced by means of the following simple experiment:

If a ray of ordinary light from a source *S*, Fig. 510, falls upon a piece of unsilvered glass at *A*, the angle of incidence being 57°, the light is divided into two portions, one of which *AB* is reflected, while the other is refracted. If the mirror is now turned so as to reflect the ray *SA* in all possible planes, without, however, changing the angle of incidence, that is, if the mirror is turned about the incident ray *SA* as an axis, it will be found that this ray may be reflected in all planes equally well, and therefore it may be inferred that the ray of light *SA* is alike on all sides. If the light in this incident ray were not alike on all sides, the mirror, which always remains under the same angle of incidence, could reflect the light more freely in one direction than in another; but since the reflected ray AB is equally bright in every plane, it follows that it can not have properties on one side different from those which it has on any other.

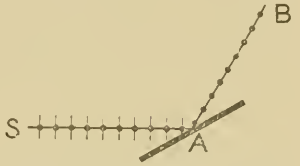

FIG. 510. A beam of ordinary light may be reflected by unsilvered glass in all directions.

If, however, Fig. 511, a second mirror is placed at *B*, so as to reflect the ray of light *AB*, the angle of incidence upon this second mirror being also 57°, it will be found that *the second mirror will reflect this ray in the plane of incidence of the first mirror, but will refuse to reflect it in a direction at right angles to this plane.* That is, as long as the second mirror is turned so that the plane of incidence *ABC* upon this mirror is the same as the plane of incidence *SAB* upon the first, light will be copiously reflected in the direction *BC;* but if the second mirror is so turned that these two planes of incidence are at right angles to each other, the light which is reflected from the first mirror will not be reflected from the second. A change has therefore taken place in the light which is reflected from the first mirror, so that the *reflected ray AB is not ordinary light,* since it may be again reflected in all directions in the plane of incidence (in the plane of the paper), but cannot be reflected at right angles to this plane. Light, like the ray *AB*, which may be reflected in one direction but not in a direction at

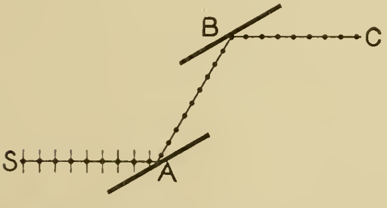

FIG. 511. A second reflection from unsilvered glass may take place in the first plane of incidence, but not at right angles to this plane.

right angles to this, has, therefore, undergone a change which makes it unsymmetrical, and this light is said to be *polarized.* A polarized beam of light is therefore a two-sided beam, having properties on one side which it does not possess upon the other.

This two-sidedness which the beam of light *AB* has acquired has a direct and most important bearing upon the question as to the direction of vibration of the ether particles which constitute this beam. If these vibrations were longitudinal, they would take place in the direction *AB*, as the light is reflected from the mirror. This beam would then be symmetrical on all sides, and it could not have properties on one side in any way different from those which it possessed on any other. It follows, therefore, that *the light AB cannot consist of longitudinal vibrations,* since, if this were so, this beam would then be symmetrical. Accordingly, *the simple fact of the two-sidedness of a beam of polarized light proves absolutely that light vibrations cannot be longitudinal, and consequently they must be transverse.* A beam of sound, however, after it has been once reflected may be reflected again in all directions, and therefore does not possess this two-sidedness. The vibrations of the air particles in a beam of sound must accordingly be longitudinal, since sound cannot be polarized even to the slightest degree.

Moreover, since the ray of light *AB* may be reflected in the direction *BC* in the plane of incidence, and since the ray *BC* may be again reflected in the plane of incidence, it follows that the light has undergone no change at the second reflection, and therefore that the vibrations of the ether particles in the ray *BC* must be in the same direction with respect to the mirror at *B* as the vibrations in the ray *AB*. The vibrations must therefore take place *parallel to the surface of the mirror at B,* that is, both in the ray *AB* and in the ray *BC the vibrations take place at right angles to the plane of incidence.* The plane *ABC*, in which polarized light may be most copiously reflected by an unsilvered mirror at the polarizing angle, is known as *the plane of polarization.* With Fresnel, then, we believe that *the direction of vibration of the ether particles in a beam of polarized light is perpendicular to the plane of polarization.*

There are many other means of polarizing light besides reflection from unsilvered glass at the polarizing angle; but however this light may be produced, the fact of its polarization may always be tested by reflection from a mirror of unsilvered glass, as at *B*. If this mirror reflects the light more freely in one direction than in another, the light is more or less completely polarized, the plane of polarization being the plane of incidence in which the light is most copiously reflected. Any device, like the mirror *A*, for producing a beam of polarized light is called a *polarizer*, and any device, like the mirror *B*, for testing the fact of polarization is called an *analyzer*.

Almost all of the light which is reflected from polished non-metallic surfaces, such as the light reflected from the surface of water, from varnished surfaces, and even from portions of the sky, is polarized. With each substance there is one particular angle of incidence for which the reflected light is more completely polarized than any other, and this angle of incidence, which differs with different substances, is known as the polarizing angle. Brewster has shown that the angle of polarization is that particular angle of incidence for which the reflected and the refracted rays make an angle of 90° with each other. An interesting relation between the polarizing angle and the index of refraction of the substance then follows:

Let *mm'*, Fig. 512, be the polished surface of a refracting medium, *AB* an incident ray of light, *BC* the reflected ray, and *BR* the refracted ray. *ABN* is the angle of incidence, and *RBN'* is the angle of refraction. Brewster found that the reflected ray *BC* is completely polarized, and therefore that the angle *ABN* is the polarizing angle, when the reflected ray *BC* and the refracted ray *BR* make an angle of 90° with each other. In this case, angle *RBN'* equals angle *m'BC*, and therefore also equals the angle *mBA*. The angle of refrac-

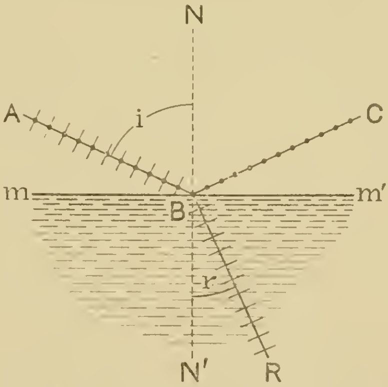

Fig. 512. Both rays are completely polarized when the reflected and the refracted rays make an angle of 90° with each other.

tion r is therefore the complement of the angle of incidence i, so that

$$\sin r = \cos i.$$

Therefore,

$$n = \frac{\sin i}{\sin r} = \frac{\sin i}{\cos i} = \tan i.$$

According to Brewster's law, therefore, *the polarizing angle for any substance is that angle of incidence the tangent of which is the index of refraction of that substance.* Investigation also shows that the refracted ray BR is likewise polarized, the plane of polarization of this ray, however, being at right angles to that of the reflected ray BC.

We are now in a position to understand the change which the ordinary light in the beam SA, Fig. 510, undergoes at A. We have learned that the vibrations of light are transverse; and therefore, as ordinary light proceeds from S to A, the vibrations take place at right angles to this direction, *but in all possible planes.* Accordingly, if we could look at the end of this beam of light, we should see these vibrations taking place in all possible directions, Fig. 513, like the spokes of a wheel. When these vibrations, which thus take place in all possible planes through the ray, reach the mirror A, *those vibrations which are already taking place parallel to the surface of the mirror are reflected, while those which are taking place at right angles to this direction are not reflected but are refracted.* Moreover, all vibrations in intermediate directions are resolved into two components, one of which is parallel to the surface of the mirror, the other being at right angles to this direction. *Those components of the vibration which are parallel to the surface of the mirror, and therefore at right angles to the plane of incidence, are reflected, while those components which are perpendicular to the first direction, and which lie therefore in the plane*

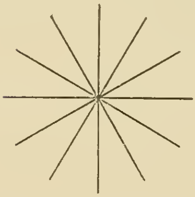

FIG. 513. In ordinary light the vibrations take place in all possible planes through the ray.

of incidence, are refracted. Accordingly, the reflected and the refracted beams are both polarized, and polarized in planes at right angles to each other.

678. Double Refraction. Another most remarkable fact closely connected with the polarization of light is the phenomenon of *double refraction.* It is found that all crystalline bodies which are founded upon a fundamental form different from that of the cube have the power of doubly refracting light, so that a single ray of light which enters the crystal emerges as two rays. Prominent among the crystals which thus have the power of doubly refracting light is the mineral Iceland spar, or calcite. If a fragment of this substance is laid upon a dot upon a piece of paper, the dot will, in general, appear double, or if a beam of light is passed through the crystal two beams will emerge, and will appear as two separate beams upon the screen.

Iceland spar crystallizes in a form known as the rhombohedron, which is a solid bounded by plane surfaces, each side of which is a rhombus. If we imagine twelve wires fastened together, forming the twelve edges of a cube, and then imagine this cube placed on one corner as *b*, Fig. 514, and pressed at the opposite diagonal corner *a*, the cube will be distorted so that three obtuse angles will meet at each of the solid angles *a* and *b*, and the cube will be deformed into a rhombohedron. The line which forms equal angles with the three edges of these solid obtuse angles, or any line parallel to this direction, is known as the *optic axis,* since the structure of the crystal is symmetrical about this direction. If the crystal is equilateral, and is therefore in the form of a true rhombohedron, as shown in the diagram, the optic axis is the line *ab*, which joins the opposite diagonal solid obtuse angles. *Every plane perpendicular to a face of the crystal and passing through the optic axis is known as a principal plane, or a principal section, of the crystal. Any plane, therefore, which contains the incident ray and the optic axis is a principal section of the crystal.*

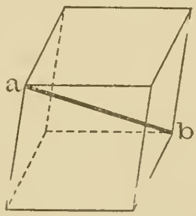

FIG. 514. A rhombohedron of Iceland spar, showing the direction of the optic axis.

If the obtuse angles at *a* and *b* are cut off by planes so that a ray of light may be sent parallel to the optic axis, it will be found that but one spot of light instead of two will appear upon the screen, from which it is evident that *no double refraction takes place when light passes parallel to the optic axis.* If a ray of light is sent through a crystal of Iceland spar in any direction other than parallel to the optic axis, Fig. 515, the single ray which enters the crystal will be divided into two rays which emerge, thus producing two images of the source upon the screen. If the crystal is now turned around the incident beam as an axis, one spot of light will be observed to remain stationary while the other revolves about it. The one which remains stationary obeys the ordinary law of refraction, and the refracted ray remains in the plane of incidence, while the ray which revolves about the first does not obey the ordinary law of refraction and the refracted ray does not always remain in the plane of incidence. The former is therefore called the *ordinary ray, O,* and the latter is called the *extraordinary ray, E.*

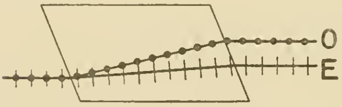

FIG. 515. Double refraction and polarization produced by Iceland spar.

Moreover, if these rays are tested for polarization by reflecting them from a piece of unsilvered glass at the polarizing angle, it will be found that *both the ordinary and the extraordinary rays are polarized, that they are polarized in planes at right angles to each other, and that the vibrations of the ordinary ray are in a plane at right angles to the principal section of the crystal, while the vibrations of the extraordinary ray are parallel to the principal section.* Measurements made upon these two rays show that *the index of refraction for the ordinary ray is constant,* and the same in all directions, *while that for the extraordinary ray is not constant,* but has different values depending upon the direction in which the ray travels within the crystal. It is evident, therefore, that *the ordinary and the extraordinary rays are propagated with different velocities within the crystal,* that the ordinary ray is propagated with equal velocity in all directions, while the extraordinary ray is propagated with different velocities in different directions. Furthermore, *since there is no double refraction in a direction parallel to the optic axis, is is also evident that the ordinary and the extraordinary rays are propagated with the same velocity in this direction.*

These phenomena show conclusively that if we could place a luminous source inside a crystal of Iceland spar, as at *C*, Fig. 516, the wave front would consist really of two wave fronts, one of which is a sphere about the luminous point as a center, and which is the wave front corresponding to the ordinary ray, while the other is an ellipsoid of revolution, also having the luminous point as a center, and which is the wave front corresponding to the extraordinary ray. Since the ordinary and the extraordinary rays travel with the same velocity in the direction of the optic axis, the sphere and the ellipsoid are tangent to each other in this direction. Every section of the wave front through the optic axis gives a circle bounded by an ellipse, with the optic axis as their common axis as revolution. It is found that certain crystals have two optic axes, so that all crystals are divided into two classes, uniaxial and biaxial. It is also found that there are two classes of uniaxial crystals, namely, those, like Iceland spar, in which the ellipse is outside the circle, which are called negative crystals, and those like quartz, Fig. 517, in which the circle is outside the ellipse, which are called positive crystals. In negative crystals, like Iceland spar, the extraordinary ray travels faster than the ordinary, and in positive crystals, such as quartz, the ordinary ray travels faster than the extraordinary.

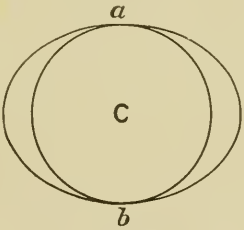

Fig. 516. The wave front in Iceland spar.

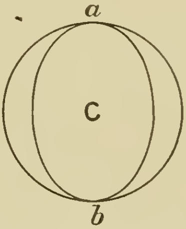

Fig. 517. The wave front in quartz.

The explanation of the phenomenon of double refraction is found in the theory offered by Fresnel, that the elasticity of the ether within the crystal is different in different directions, and hence the velocity with which light is propagated must likewise vary with the direction. There is, moreover, much experimental evidence in support of this theory. If a rod is sawed from a crystal of a doubly refracting substance and one end is clamped in a vise, the other end may be bowed like a tuning fork and so be made to emit a musical note. It is found, however, that the note which is thus emitted by the vibrating rod depends upon the direction in which the latter is bowed, thus proving that in one direction the substance is more elastic even to mechanical disturbances than in a direction at right angles to the first. Moreover, by a beautiful method due to Professor Kundt, known as the "breath figures", it may be shown that the conductivity for heat of a doubly refracting substance is greater in one direction than in a direction at right angles to this. If a crystal of quartz or one of Iceland spar is covered with moisture from the breath, and is then touched by a heated point, such as the tang of a file, the crystal immediately about the point will be warmed, and the moisture will be evaporated. If, now, lycopodium powder is dusted over the entire surface, the powder will adhere to the part that is still moist, but may be blown away from that which is dry. It will then be found, Fig. 518, that the moisture has been evaporated from an area which is elliptical in shape, thereby showing that heat has been conducted by the crystal more rapidly in the direction of the major axis than in that of the minor axis of this ellipse. Hence we may believe, and experiment shows this to be true, that the optical elasticity of the crystal is likewise greatest in the direction in which the conductivity of the crystal for heat is a maximum.

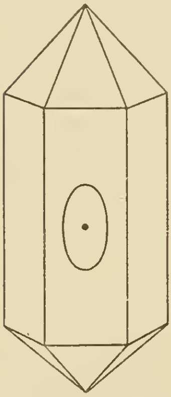

FIG. 518. Quartz conducts heat most readily in a direction parallel to its optic axis.

From what has been said, it is evident that the vibrations of the ether within a doubly refracting crystal take place in two planes only, one being a principal plane, and the other a plane at right angles to this. Let these planes be denoted by *E* and *O* respectively. If a ray of polarized light, *PP*, Fig. 519, enters the crystal, the vibrations of this ray making an angle with *E* and *O*, these vibrations are resolved into two components, the magnitudes of which depend upon the inclination of the vibrations to *E* and *O*. If the original vibrations make an angle of 45° with *E* and *O*, both of these components will be equal. If the original vibrations lie in the plane *O*, the component in *E* will be zero; and if the original vibrations lie in the plane *E*, the component in *O* will be zero.

Accordingly, in general, if a beam of polarized light is passed through a piece of Iceland spar, the vibrations of the beam within the spar will be divided into two components and these will appear as two spots of light upon the screen. If the crystal is rotated, the brilliancy of one of these components will increase while that of the other will decrease, until the principal section of the crystal is parallel to the original vibrations, and then only the extraordinary beam passes, and but one spot of light appears upon the screen. As the crystal is further rotated, the ordinary beam reappears and grows brighter, while the extraordinary beam grows fainter, until the crystal is so turned that its principal section is at right angles to the original direction of the vibrations, and then only the ordinary beam is transmitted, while the extraordinary ray is extinguished. Upon continued rotation of the crystal, this process is repeated indefinitely, the beam which is extinguished in one position of the crystal appearing brightest when the rotation is carried 90° farther.

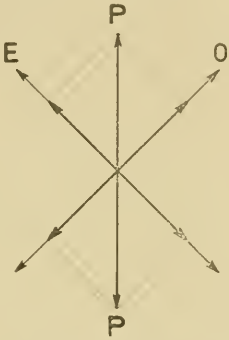

FIG. 519. Components into which polarized light is resolved within a doubly refracting crystal.

If light which has been divided into an ordinary and an extraordinary beam by one piece of Iceland spar is allowed to fall upon a second similar piece of spar, each of the two beams emerging from the first crystal will undergo resolution into two components upon entering the second, so that, in general, there will be four beams upon the screen. When, however, the principal sections of the two crystals are parallel, the light which passes through the first vibrating in two directions, will pass through the second vibrating in the same two directions, and in this position of the crystals there will appear but two beams. Also, when the second crystal is turned 90° from this position, the ordinary and the extraordinary rays which pass through the first crystal will now pass as extraordinary and ordinary through the second, and hence there will again be only two beams of light upon the screen. When the crystal is rotated still farther so that the planes are again parallel but 180° from their first position, the separation into two beams produced by the first crystal is counterbalanced by an equal separation in the opposite direction produced by the second, and but a single spot of light appears.

679. The Nicol's Prism. Inasmuch as Iceland spar is perfectly transparent, and polarizes as well as doubly refracts the light which passes through it, a crystal of this substance would make an effective means of producing a beam of polarized light if only it were possible to do away with one or the other of the two beams which are always produced. If the angular separation of these beams were greater, this could be accomplished by placing a little screen so as to intercept either one or the other. Nicol, however, accomplished this same purpose in an even more ingenious manner. He conceived the idea of introducing into the crystal a layer of a substance upon which the two rays would be incident at such angles that one would be transmitted while the other would undergo total reflection. A crystal about three times as great in length as in diameter was then sawed into two parts along the line *AB* joining the two obtuse angles, Fig. 520. The two cut surfaces were then polished and cemented together in their former positions with Canada balsam, so that the crystal was once more in its original condition, with the exception that it contained a diagonal sheet of this transparent cementing material. The natural end surfaces of the crystal were also slightly changed in inclination in order to obtain the proper angle of incidence upon this totally reflecting layer.

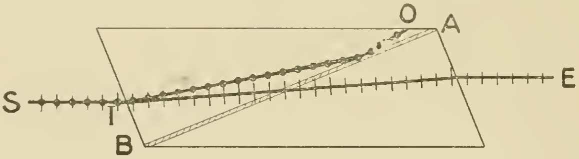

FIG. 520. A Nicol's prism transmits only the extraordinary ray, the ordinary ray being totally reflected by the Canada balsam.

The index of refraction of Canada balsam is less than that of the spar for the ordinary ray and greater than that of the spar for the extraordinary ray. In other words, the Canada balsam is a rarer medium than the spar of the ordinary ray and a denser medium than the spar for the extraordinary ray. Accordingly, when a ray of ordinary light *SI* is incident upon a crystal of Iceland spar which has been so treated, it divides into an ordinary and an extraordinary ray. *The ordinary ray*, which is incident upon the the layer of Canada balsam at an angle greater than the critical angle, *then undergoes total reflection, and is reflected out to one side of the prism*, where it is absorbed by black paint. *The extraordinary ray*, on the other hand, *passes on through the Canada balsam and so through the prism*, thus producing a beautifully brilliant beam of light which is completely polarized.

A crystal of Iceland spar which has been so treated and, which thus produces a single beam of polarized light, is known as a Nicol's prism, or, more simply, as a Nicol. A Nicol's prism is the most convenient as well as the most desirable means of producing a beam of polarized light. *Since it is the extraordinary ray which is thus transmitted*, the vibrations of the light which passes through a Nicol's prism take place *in the principal plane of the crystal*, that is parallel to the shorter diagonal of the diamond-shaped figure which the end of the Nicol presents, Fig. 521. A second Nicol's prism similar to the first serves also most excellently as an analyzer.

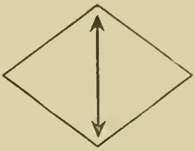

FIG. 521. Direction of vibration of the light in a Nicol's prism.

If the principal plane of the second Nicol, or the analyzer, is parallel to that of the first, or the polarizer, the light which passes through the polarizer will also pass through the analyzer, and light will be transmitted to the screen. But if the principal plane of the analyzer makes an angle with that of the polarizer, the light from the polarizer is resolved into two components, thus forming an ordinary and an extraordinary ray in the analyzer, one of which, the ordinary ray, when it meets the Canada balsam, is totally reflected out to the side of the prism and is absorbed, while the other, the extraordinary ray, passes through. The component which is thus transmitted by the analyzer becomes less as the angle between the principal planes of the two Nicols increases, and when the principal plane of the analyzer makes an angle of 90° with the principal plane of the polarizer, the entire extraordinary ray from the polarizer enters the analyzer as an ordinary ray, and so is totally reflected. Therefore, when the principal planes of the polarizer and the analyzer make an angle of 90° with each other, the light which passes the polarizer is wholly excluded by the analyzer, and no light is transmitted to the screen. The polarizer and analyzer are then said to be *crossed.* As the analyzer is rotated still farther, a component of the light from the polarizer again passes through the analyzer, and this transmitted component increases in brightness until the two planes are again *parallel,* when the light once more attains its maximum intensity.

680. The Polarization Produced by Tourmaline. The mineral tourmaline is a doubly refracting surface, and, like all doubly refracting media, it polarizes as well as doubly refracts the light which passes through it. This mineral, however, possesses the peculiar property of absorbing the ordinary ray, while it transmits freely the extraordinary ray, so that a crystal of tourmaline, when placed in the path of a beam of ordinary light, will transmit a beam which is completely polarized, since it consists entirely of the extraordinary ray. Indeed, if it were not for the fact that this material is colored, the tint being usually a deep brown or a light green, a plate of this substance could be most conveniently used as a means of polarizing the light.

The cause of the absorption of the ordinary ray in tourmaline is in itself most interesting. Inasmuch as this substance is a *crystal,* it follows that its elasticity in one direction is greater than that in a direction at right angles to this. Accordingly, the free periods in which its particles vibrate are different in different directions. In the direction within the crystal in which the vibrations of the ordinary ray take place, it chances that the natural period of the tourmaline particles is practically the same as that of the light vibrations, so that the latter are absorbed sympathetically. In a direction at right angles to this, however, that is, in the direction within the crystal in which the vibrations of the extraordinary ray occur, the natural period of the particles is entirely different from that of the light vibrations, hence the latter are not thus absorbed, and the extraordinary ray passes through the crystal, producing a single beam of polarized light.

Accordingly, two pieces of tourmaline, whether from the same crystal or from different crystals, serve admirably the purpose of polarizer and analyzer, and, except for the objection of the deep color which is always present, these may be used in every way instead of two Nicol's prisms. If the second crystal is placed with its principal plane parallel to that of the first, as at *AB*, Fig. 522, the light which is transmitted by the first passes freely through the second; but if the crystal is turned so that its principal plane is at right angles to that of the first, as at *A″B″*, the light which passes the first as an extraordinary ray now enters the second as an ordinary ray and is absorbed, so that no light is transmitted. If the principal plane of the second makes an angle with the first, as shown at *A′B′*, the light from the first is resolved into two components in the second, one of which is transmitted while the other is absorbed, the transmitted component becoming less as the two principal planes depart more and more from parallelism. As the second crystal is rotated from its initial position, the transmitted light diminishes in intensity until the principal planes are at right angles to each other, when all light is excluded; and as the rotation is continued, the light again appears and increases in brilliancy until the two planes are once more parallel.

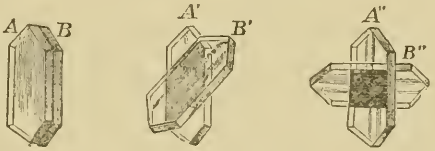

FIG. 522. The action of two tourmaline plates in transmitting light.

681. The Rotation of the Plane of Polarization. When a beam of polarized light is sent through a plate of quartz which has been cut perpendicularly to the optic axis, that is, when light passes in this medium *parallel to the optic axis,* no double refraction is produced, but the remarkable phenomenon of *the rotation of the plane of polarization* takes place. Therefore, if polarized light which is vibrating vertically, for instance, enters a plate of quartz so as to pass parallel to its optic axis, the light will emerge from the quartz still polarized, although the vibrations are no longer taking place vertically, but are inclined to their former direction. In other words, *the plane of polarization has been rotated through a certain angle.* Moreover, the angle through which this plane has been turned is different for different wave lengths, the amount of this rotation increasing as the wave length of the light becomes shorter.

Accordingly, if the vibrations of polarized white light are supposed to be vertical for all wave lengths, Fig. 523, and this polarized light, *OP*, is passed through a plate of quartz parallel to its optic basis, the light will emerge from the quartz with each wave length executing vibrations in a plane which is different from that of any other, the vibrations of the red light, for instance, making an angle of 10° with the vertical, the vibrations of the yellow light making an angle of 20° with the vertical, and so on. Although lights of different colors are executing vibrations in different planes, the resulting light appears white, since the eye receives the impression of these various colors superposed.

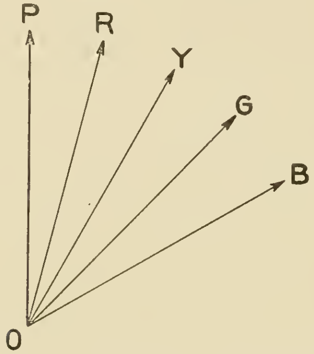

Fig. 523. The rotation of the plane of polarization is inversely proportional to the wave length.

If, however, this light is now made to pass through an analyzer, the result will be entirely different. If any vibration happens to coincide in direction with the principal plane of the analyzer, that particular color will be transmitted; and if any vibration comes to the analyzer at right angles to its principal plane, that color will be excluded. Thus, to refer to the above illustration, if the principal plane of the analyzer makes an angle of 10° with the vertical, the red light will be transmitted, since the vibrations of this particular colored light make an angle of 10° with the vertical; and if the principal plane of the analyzer makes an angle of 20° with the vertical, the yellow light will be transmitted, and so on.

In general, however, the vibrations of any wave length of light as they come to the analyzer will undergo resolution into two components, one of which lies in the principal plane of the analyzer, *A*, Fig 524, and the other in a plane at right angles to this direction. The first of these components will pass freely, since the analyzer transmits vibrations taking place in its principal plane, while the second, which is vibrating at right angles to this plane, will be unable to pass. Moreover, since the vibrations of the various colored lights make a different angle with the principal plane of the analyzer, the magnitude of the components of the light which the analyzer transmits will be different for each different color, and the light which is transmitted to the screen will be a mixture of a great many colored lights in different proportions.

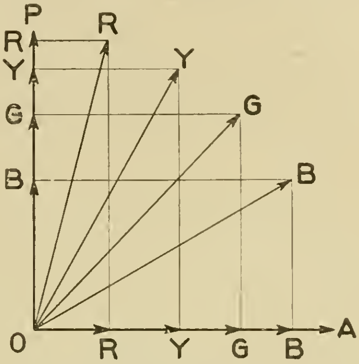

FIG. 524. One component of each vibration passes through the analyzer.

Accordingly, the light which reaches the screen will be colored, the particular color depending upon the position of the analyzer. As the analyzer is rotated, the color of the transmitted light changes, since the components of the different wave lengths which the analyzer transmits are now different and the tint of the transmitted light is always a mixture of those colors, or the components of those colors, which pass through the analyzer. When the analyzer is turned 90°, the color on the screen is absolutely complementary to the one seen in its first position, inasmuch as the analyzer in one position transmits those components which it excludes when it is turned 90°.

It follows, therefore, that if a piece of Iceland spar instead of a Nicol's prism is used as an analyzer, the light which passes through the quartz is divided into two components and therefore into two beams as it enters the spar, each component having those portions of the original white light which are not found in the other. There are thus produced upon the screen two beams of light each of which is colored, the color of one being absolutely complementary to that of the other. As the crystal of Iceland spar forming the analyzer is rotated, the proportion of each wave length of light in each component changes, and hence the color of each beam changes; but whatever the color of one may be, the color of the other is always complementary to it. If, therefore, the two beams are made partially to coincide upon the screen, the overlapping portion is white, since the colors which are present in one are lacking in the other, and hence the union of the two supplies all the colors which make up white light. This is by far the best method of showing complementary colors, and may be used instead of the yellow and the blue glasses in a mixture of colored lights, as described on page 842.

682. The Effect of Interposing a Plate of a Doubly Refracting Substance Between the Crossed Polarizer and Analyzer. If a plate of a doubly refracting substance, such as mica or selenite, is placed between the crossed polarizer and analyzer, light, in general, will be restored. This restored light will be most brilliant when the principal plane of the doubly refracting substance makes an angle of 45° with the principal plane of the polarizer, but no effect will be produced if this angle is zero or 90°. The vibrations of the light from the polarizer, as they enter the plate of mica, are resolved into two components, one vibrating in the principal plane of the mica and the other at right angles to this direction. In other words, one of these components travels through the mica as an ordinary beam and the other as an extraordinary beam. The separation of the two, however, is so very slight, particularly if the mica is thin, that what we see is only one emergent beam, but in this the vibrations are taking place in two directions at right angles to each other.

If we assume that the light which comes from the polarizer is vibrating in a vertical plane, the analyzer will then transmit only light which is vibrating in a horizontal plane, since the polarizer and analyzer are crossed. Accordingly, light from the polarizer, vibrating in the direction *PP*, Fig. 525, comes to the plate of mica, and this allows light to pass which is vibrating in the directions *ab* and *cd*. The vibration *PP* will be resolved into two components *OO* and *EE* at right angles to each other, and therefore the light which passes through the plate of mica is vibrating in these two directions. The light, still vibrating in these two directions, now comes to the analyzer, which allows, in this case, only light to pass which is vibrating horizontally.

Each of the component vibrations *OO* and *EE*, Fig 526, is now resolved into two components, one in the plane of the analyzer *AA* and the other at right angles to this direction. Those components of *OO* and *EE* which are in the plane of the analyzer can now pass through, while those components at right angles to this direction are excluded. In general, therefore, light is restored when a plate of a doubly refracting substance is placed between the crossed polarizer and analyzer.

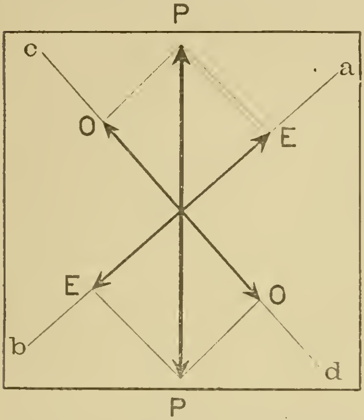

FIG. 525. Vertically vibrating light from the polarizer is resolved into two components in the mica.

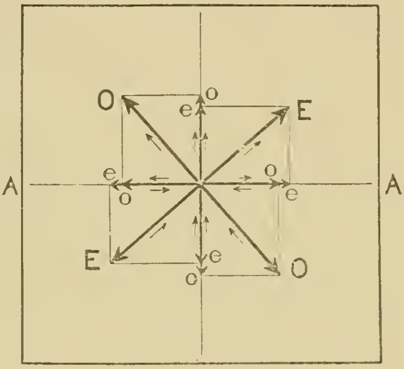

FIG. 526. Each vibration transmitted by the mica is resolved into two components at the analyzer.

Moreover, if the plate of the interposed substance is thin, the light which is thus restored is colored, as may be seen from the following examination of the vibrations of the rays. When the light from the polarizer enters the mica it is resolved into two components one of which travels through the mica as an ordinary ray and the other as an extraordinary ray, and therefore one ray traverses the mica a little faster than the other. When the two beams emerge, there is a difference of phase between them, since one ray has been retarded more than the other in its passage through the mica. It would seem at first sight, therefore, that these two emergent beams were capable of destroying each other by interference, since a difference of phase exists between them. It has been found, however that *two beams of light can never interfere unless they are executing vibrations in the same plane.* Inasmuch, therefore, as these two beams of light are vibrating at right angles to each other, they cannot interfere, and therefore they both come to the analyzer without mutually destroying each other.

When each of these two beams is resolved into two components by the analyzer, it will be seen that one component *o* of the ordinary beam passes through the analyzer and one component *o* is excluded, and also that one component *e* of the extraordinary beam passes through the analyzer while the other component *e* is excluded. The light which now passes through the analyzer consists partly of the light which passed through the mica as an ordinary ray and partly of that which passed as an extraordinary ray. Moreover, since one of these components has travelled faster in the mica than the other, a *difference of phase exists between the two components which pass through the analyzer; and furthermore, since these two components are now vibrating in the same plane, they are capable of interfering.* This difference of phase will be exactly half a wave length for light of some one color, and hence that particular color will be destroyed by interference, and the color of the light which reaches the screen is a mixture of all the other colors which are not thus destroyed. Since the amount by which one ray is retarded more than the other depends upon the thickness of the mica, the tint of the color which is transmitted by the analyzer will depend upon the thickness of the plate through which the light has passed. If the plate of mica is of different thicknesses, as is usually the case when films are flaked off, a different color for each different thickness of the film will be transmitted to the screen.

If the analyzer is now rotated, the colors become dim; and when the principal plane of the analyzer coincides in direction with either *OO* or *EE*, the restored light is white. In other words, if the plane of the analyzer is parallel to the direction *EE* the vibrations in this direction as they come from the mica can pass through while those in the direction *OO* are excluded; and if the plane of the analyzer is parallel to the direction *OO*, then it is the vibrations in this direction which can now pass through the analyzer, while those in the direction *EE*, in their turn are excluded. In either case, no interference takes place, hence no color is destroyed, and the transmitted light is white. If the analyzer is rotated until its principal plane is 90° from its first position, the restored light is again colored, and the colors are complementary to those which were transmitted previously. The production of complementary colors in the two positions of the analyzer at 90° from each other may be explained with the aid of Fig. 526.

If we assume, for the moment, that a particular vibration, as it comes from the polarizer, is taking place vertically upward, the two components O and E of this vibration, as they come through the mica, are vibrating diagonally upward. Moreover, each of these components itself undergoes resolution into two components at the analyzer, one of the latter components being vertical and the other horizontal. It is thus seen, therefore, not only that there are two components, an o and an e, in the vertical direction, and two components, an o and an e, in the horizontal direction, but also that *these two components are in the same phase in the vertical direction and in opposite phase in the horizontal direction.* The o and the e, however, both in the vertical and in the horizontal directions, have a certain difference of phase between them due to the fact that one has travelled in the mica faster than the other. Accordingly, if the plane of the analyzer is vertical, these vertical components o and e pass through, and destructive interference will occur for that particular color for which the difference of phase between them is half a wave length.

Furthermore, when the analyzer is turned into the horizontal position, the horizontal components o and e are now transmitted and destructive interference will again be produced. In the case of these components, however, a difference of phase of half a *wave length, due to the difference in the direction of vibration, has been added to the difference of phase already existing between them due to the difference in their rate of travel through the mica.* From this it follows that the components o and e in the horizontal direction and the same components in the vertical direction are vibrating in *exactly opposite phase*, so that the particular color which was transmitted in the vertical direction will now be destroyed, and the color which was destroyed in the vertical direction will now be transmitted. In other words, the analyzer in one position transmits all those wave lengths which it excludes in a position at right angles to this, and hence the tint produced by the mixture of the colors which the analyzer transmits in one position is absolutely complementary to that produced by the mixture of the colors which the analyzer allows to pass after being rotated through an angle of 90°.

If plates of selenite are ground to different thicknesses, thereby forming patterns of different designs, such as a butterfly, a tulip, or a rose, and these plates are interposed between the crossed polarizer and analyzer, the light which is restored will cause these objects to appear in their natural colors, since the plates have been ground to such a depth as to give the required color by interference. When a piece of glass is heated, or compressed or bent, it acquires the power of doubly refracting the light, and hence becomes temporarily crystalline. Glass which has been suddenly cooled without being annealed also has the power of doubly refracting the light, so that if such glasses are thus examined by polarized light, they will be seen to possess a strained structure which is very complex. It is therefore important to examine in this way all glass which is to be used in the construction of optical instruments in order to see that it is free from tension or distortion, for if strains or lack of annealing were to exist, this glass would doubly refract the light which passes through it.

683. The Rings and Cross in Convergent Polarized Light. In the preceding paragraph we have seen that light may be restored when a plate of a doubly refracting substance is placed between the crossed polarizer and analyzer. In obtaining these results the crystal was placed in *parallel* polarized light, and the plate was cut *parallel to the optic axis,* so that the light passed through the crystal at right angles to this direction.

Phenomena of an entirely different nature are obtained when the light from the polarizer is made to pass through a lens of short focal length so as to be *convergent,* and when a plate of a doubly refracting substance cut *perpendicularly* to the optic axis is placed in this convergent beam of polarized light, Fig. 527. The axis of this conical beam then coincides in direction with the optic axis of the substance, and therefore passes through the plate without undergoing double refraction. All other rays, however, that is, all those which pass through the crystal in an inclined direction, will be doubly refracted. We have learned, moreover, that when light passes through a plate of a doubly refracting medium, the vibrations are resolved into two components, one taking place in the principal section of the substance and the other in a plane at right angles to this direction. Inasmuch as the the principal section of a crystal is the plane containing the ray and the optic axis, and inasmuch as the optic axis is here normal to the surface of the plate, it follows that *every plane of incidence is a principal section.*

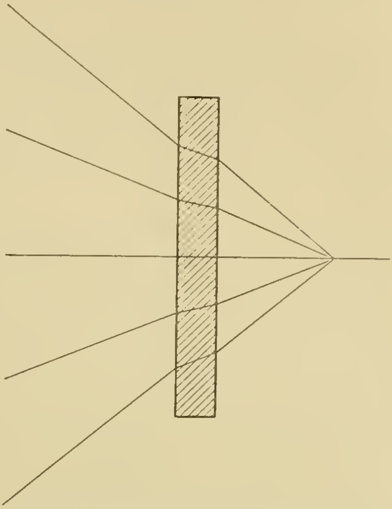

FIG. 527. Convergent polarized light passing through a doubly refracting substance cut perpendicularly to the optic axis.

Therefore, if the plate of crystal coincides with the plane of the paper, Fig. 528, and if the axis of the conical beam enters the plate normal to the surface at C', *every plane of incidence passing through C' is a principal section,* since the axis of the conical beam coincides with the direction of the optic axis of the crystal. Accordingly, if a ray of light is incident at any point of the crystal, as at A, B, C, or D, the straight line connecting this point with the axis C' is a principal section. The light therefore, which is doubly refracted passes through the crystal as an extraordinary ray vibrating in this direction, and as an ordinary ray vibrating at right angles to this direction. If the light

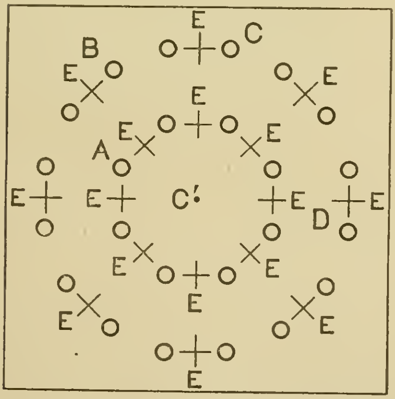

FIG. 528. Every plane of incidence passing through C' is a principal section of the crystal.

from the polarizer is vibrating vertically, and if a ray of polarized light from this conical pencil is incident upon the plate at *A*, for instance, this ray of light will be resolved into two components *O* and *E* which will travel through the plate in an inclined direction with different velocities. When this light, now vibrating in these two directions, comes to the analyzer, each of these rays, the *O* and the *E*, will be resolved into two components, one in the principal plane of the analyzer, which will be transmitted, and the other at right angles to this plane, which will be excluded.

Since the two rays *O* and *E* travelled through the plate with different velocities, the components of each of these rays which pass through the analyzer will be exactly opposite in phase for some one color, hence this color will be destroyed by interference. Therefore, the light which is restored will be colored, this color being a mixture of all the other colors which compose white light except the one thus destroyed. Since the color depends upon the difference of phase between the two components in the analyzer, and since this difference of phase, in turn, depends upon the relative retardation of the rays *O* and *E* as they traverse the plate, that is, upon the obliquity of the light, *any particular color will be uniform in a ring around the axis of the conical beam as a center.* Moreover, if a ray is incident upon the plate at a greater obliquity, as at *B*, the relative retardation between *O* and *E* will be greater on account of the greater distance travelled in the crystal, and the difference of phase between the two components in the analyzer will be exactly half a wave length for some other color. Some other color will then be destroyed by interference, and the color transmitted will be different. The light which is restored consists, therefore, *of a series of concentric colored rings about the axis of the conical beam.*

Superposed upon this system of colored rings, however, appears a cross, this cross being *black* when the principal planes of the polarizer and analyzer are *crossed,* and *white* when these planes are *parallel* to each other. The explanation of this cross, as well as the change from black to white which it undergoes, is simple. Remembering that the vibrations of the light are vertical as they come from the polarizer, we see that if a ray of this vertically vibrating polarized light is incident upon the plate at

C, it will pass straight on through the plate as an extraordinary ray without double refraction, since at this point the principal plane of the plate is parallel to the direction of vibration of the light from the polarizer. Similarly, the ray which meets the plate at *D* will pass through as an ordinary ray, likewise without double refraction. Therefore, in a vertical and also in a horizontal plane, the vertically vibrating light which falls upon this plate passes on through and comes to the analyzer still vibrating vertically.

If, now, the planes of the polarizer and analyzer are at right angles to each other, the analyzer can transmit only light which is vibrating horizontally, and therefore *it will exclude the light in these two directions.* A black cross superposed upon the system of rings will accordingly appear, Fig. 529, the arms of the cross coinciding with the principal planes of the crossed polarizer and analyzer. If, however, the analyzer is rotated through 90° so that its principal plane is parallel to that of the polarizer, the light which is vibrating vertically in the vertical and in the horizontal planes can now pass, and the black cross will be replaced by a white one, Fig. 530. Accordingly, if a plate of a uniaxial crystal, cut perpendicularly to the optic axis, is placed in convergent light between the polarizer and analyzer, *a black cross superposed upon a system of colored rings will appear if the polarizer and analyzer are crossed, and a white cross super-*

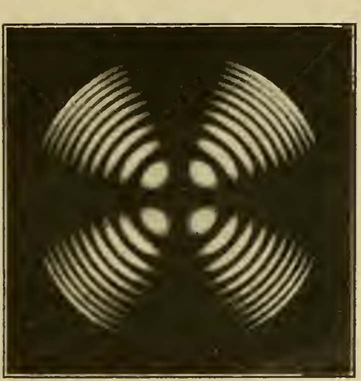

Fig. 529. A uniaxial crystal in convergent polarized light—the polarizer and analyzer crossed.

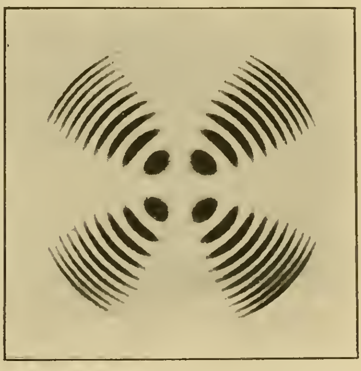

Fig. 530. A uniaxial crystal in convergent polarized light—the polarizer and analyzer parallel.

posed upon a system of rings of the complementary colors will be seen if the planes of the polarizer and analyzer are parallel.

The phenomena observed in the case of a biaxial crystal are far more complex. If a biaxial crystal is cut by a plane perpendicular to the line midway between the two optic axes, so that convergent light can be sent through the plate in a direction parallel to this line, a system of colored rings will encircle each axis, and these, in turn, will be surrounded by a leminscate and a series of ovals. If the line joining the optic axes of the plate coincides with the principal plane of the polarizer or the analyzer, a dark cross-shaped figure will be superposed upon this system of rings and ovals, Fig. 531; but if this line makes an angle of 45° with the planes of the polarizer and analyzer, this series of colored bands will be crossed by two hyperbola-shaped black curves passing through the optic axes, Fig. 532. All these phenomena are complementary when the analyzer is turned throught 90°. It is interesting to notice that many of these phenomena, although so very complex, were predicted by a mathematical investigation based upon the wave theory of light before they were tried experimentally.

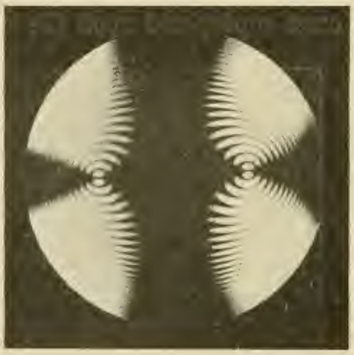

FIG. 531. A biaxial crystal in convergent polarized light—the line joining the optic axes being parallel to the principal plane of the polarizer.

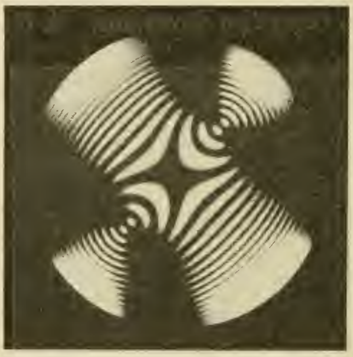

FIG. 532. A biaxial crystal in convergent polarized light—the line joining the optic axes being inclined 45° to the principal plane of the polarizer.

Since the action of heat causes a crystal to expand unequally in the directions of its different elasticities, is not impossible to conceive of the application of heat as so affecting the elasticity of the ether within the crystal as even to change its crystalline structure. At ordinary temperatures the mineral selenite is a biaxial crystal, and the intricate curves peculiar to biaxial crystals are obtained when this substance is examined in convergent polarized light. When a plate of this material is gently heated, however, the elasticity of the ether within the crystal is so changed that the two axes are seen to come nearer together, the lemniscate and ovals approach the circular form and surround the axes, and the dark hyperbola-shaped curves move towards each other. This continues, until, for a moment, when the two axes coincide, the substance has every property of a uniaxial crystal, the complicated curves then become true circles about the single axis, and the branches of the hyperbola-shaped figures unite to form the cross. The two axes are then seen to draw apart in the opposite directions, and the crystal is once more biaxial. As the mineral cools, these processes all take place in the reverse order, the whole forming a phenomenon of surpassing interest and beauty.

Evolution is the [illegible] of change

stone Bridge 5/3 p